Volume 25 Number 1 March 1974
BJPIA5 25 (No. 1) 1-97 (1974) 0007-0882

The British Journal for the Philosophy of Science

16

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Acting Editor: JOHN WORRALL

Assistant Editors: COLIN HOWSON AND ELIE ZAHAR

Subscriptions

The *Journal* is published quarterly. The subscription price is £7.00 net (U.S.A. and Canada U.S. $22.00) per volume post free; single issues £2.50 net (U.S.A. and Canada U.S. $8.00) plus postage. The subscription is £4.50 (U.S.A. and Canada U.S. $14.00) for members of the British Society for the Philosophy of Science, who should order direct from the publisher. (Membership of the Society is international and is by no means confined to people resident in the United Kingdom.) Orders may be sent to any bookseller or direct to Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT.

Notes to Contributors

Articles, correspondence and books for review should be sent to the Acting Editor, John Worrall at the Department of Philosophy, London School of Economics, Houghton Street, Aldwych, London WC2A 2AE. Articles should be accompanied by abstracts of about 100 words.

Material is accepted for publication in the *Journal* on the condition that copyright is assigned to the British Society for the Philosophy of Science.

Three copies of all contributions should be sent, typed on one side only with wide margins and double-spaced. Footnotes, which should be as few as possible, should be typed separately, double-spaced at the end of the text. They should be referred to by consecutive numbering throughout the text, in which they will be inserted in page proof. Bibliographical references should be listed alphabetically at the end. The style of references should follow current *Journal* practice.

Long quotations are printed in small type without quotes, and should be so marked in the typescript. Single quotes should be used to form the names of words and sentences, and double quotes for all other purposes. Diagrams, tables, and illustrations should be on separate sheets, with their desired position in the text clearly indicated. Symbolic formulae appearing as separate lines should be clearly distinguished in the text, and foreign letters (Greek, German, etc.) indicated in the margin. Material to be printed in italics, especially English letters used as symbols ('*A*', '*x*', etc.), should be underlined. On matters of style and layout, authors should consult CHAUNDY, T. W. *et al.* [1964]: *The Printing of Mathematics*, London: Oxford University Press.

Offprints of articles may be ordered when corrected galley proofs are returned. Fifty are supplied free ; the price of extra copies is given on the order form.

© British Society for the Philosophy of Science 1974

For permission to reproduce material from *The British Journal for the Philosophy of Science*, please apply to the Editor.

P 60

Back Volumes

Volumes 1-16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17-24 may be obtained from Cambridge University Press.

CU/C-H00665-16-P860 501.05
B774

Brit. J. Phil. Sci. 25 (1974), 1–23 *Printed in Great Britain*

Logical versus Historical Theories of Confirmation

by ALAN MUSGRAVE

1 *Introduction.*
2 *'Background Knowledge' and the Paradox of Confirmation.*
3 *The Strictly Temporal View of Background Knowledge.*
4 *The Heuristic View of Background Knowledge.*
5 *Background Knowledge or a Background Theory?*
6 *Requirements for the Growth of Knowledge.*
7 *Conclusion.*

1 INTRODUCTION

Thales predicted an eclipse, and became one of the Seven Sages. Since then many have urged that a scientific theory is to be especially prized if it yields successful predictions. For example, Clavius argued that Ptolemaic astronomy must be true because by using its hypotheses 'not only are all the appearances already known accounted for, but also future phenomena are predicted'.[1] Descartes rejected Ptolemaic astronomy, but accepted the methodological principle: we know that our hypotheses are correct, he says, 'only when we see that we can explain in terms of them, not merely the effects we had originally in mind, but also all other phenomena of which we did not previously think'.[2] For Leibniz it was 'the greatest commendation of an hypothesis (next to truth) if by its help predictions can be made even about phenomena not yet tried'.[3] Whewell insisted that good

[1] Christopher Clavius, *In Sphaeram Ionnis de Sacro Bosco commentarius* (1602), pp. 518–9; cited from Blake [1960], p. 34. Clavius claimed that the success of Ptolemaic *predictions* was a powerful argument for the reality of Ptolemy's eccentrics and epicycles and against an instrumentalist interpretation of them as 'geometric fictions': 'For it is incredible that we force the heavens (but we seem to force them, if the Eccentrics and Epicycles are figments, as our adversaries will have it) to obey the figments of our own minds, and to move as we will, or in accordance with our principles.' Later it was successful predictions which led Duhem to water down his instrumentalism with the idea of 'natural classification'.

[2] René Descartes, *Principles of Philosophy* (1644), Part *III*, *xlii*.

[3] Leibniz, Letter to Conring (1678); cited from Lakatos [1970], p. 123.

hypotheses 'ought to *foretell* phenomena which have not yet been observed'.[1] Duhem agreed.[2]

Quotations like this could be multiplied, from both scientists and philosophers of science. The intuitions they express are that a good hypothesis must not only account for known phenomena, but must successfully predict new ones; that explaining known facts is one thing, predicting new facts another; and that in assessing the confirmation or evidential support of a hypothesis, we must take into account especially (perhaps even exclusively) the success or failure of its *predictions*.

These intuitions have usually received a dusty response from philosophers interested in the theoretical analysis of evidential support or confirmation. According to modern logical empiricist orthodoxy, in deciding whether hypothesis *h* is confirmed by evidence *e*, and how well it is confirmed, we must consider only the statements *h* and *e*, and the logical relations between them. It is quite irrelevant whether *e* was known first and *h* proposed to explain it, or whether *e* resulted from testing predictions drawn from *h*. I shall call this cornerstone of logical empiricism the *purely logical* (or *logical*, for short) *approach to confirmation.*

We find it in Mill, who was amazed at Whewell's view that 'an hypothesis . . . is entitled to a more favourable reception, if besides accounting for all the facts previously known, it has led to the anticipation and prediction of others which experience afterwards verified. Such predictions and their fulfilment are, indeed, well calculated to impress the uninformed . . . But it is strange that any considerable stress should be laid upon such a coincidence by persons of scientific attainments'.[3] Keynes agreed, and insisted that the 'peculiar value of prediction . . . is altogether imaginary . . . The question as to whether a particular hypothesis happens to be propounded before or after examination of [its instances] is quite irrelevant'.[4] Hempel tries to answer the qualitative question of whether or not *e* confirms *h* by considering only the logical forms of *e* and *h*.[5] And Carnap's quantitative degrees of confirmation are also to depend entirely upon the logical forms of the two sentences concerned. On this view, so far as problems of confirmation or evidential support are concerned 'it is quite irrelevant whether in fact scientists always or usually or never make their observations before conceiving their theories or vice versa'.[1]

---

[1] William Whewell [1840], volume *II*, chapter *5*, p. 62. Cf. Medawar [1967], p. 149, and Lakatos [1970], p. 123.

[2] See section *2* of chapter *5* of Duhem's [1954], entitled 'Theory anticipating experiment'.

[3] J. S. Mill, *System of Logic*, *III*, *xiv*, 6. *Cf.* Medawar [1967], p. 141, and Lakatos [1970], p. 123. Mill's main point (that successful prediction cannot conclusively prove the theory involved) was, of course, quite right—and Clavius, Descartes and Whewell were wrong.

[4] J. M. Keynes [1921], p. 305. Cf. Popper [1963], pp. 247–8, and Lakatos [1970], p. 123.

[5] Hempel says: '. . . it seems reasonable to require that the criteria of empirical confirmation . . . should contain no reference to the specific subject matter of the hypothesis or of the evidence in question; it ought to be possible, one feels, to set up purely formal criteria of confirmation in a manner similar to that in which deductive logic provides purely formal criteria for the validity of deductive inference' (Hempel [1965], p. 10).

So we see that those who follow the logical approach to confirmation show scant respect for the intuition that successful prediction matters. After all, they argue, scientists use evidence to evaluate theories by checking whether their consequences are true. If the evidence shows that some consequence of a theory is true, then this cannot depend upon whether the evidence came to be known before the theory was proposed or afterwards. Such historical considerations, interesting though they may be, ought not to affect questions of confirmation or evidential support. For these questions are *objective* ones, to be settled solely by inspecting the logical relationship between theory and evidence.

Yet the purely logical approach to confirmation has run into some notorious difficulties—the most notorious of all being the *paradox of confirmation*. And the best solution to that paradox, or so it seems to me, points towards a different approach to confirmation, and one which does more justice to the intuition that successful predictions are important. It does so, as Lakatos was the first to emphasise, by introducing an *historical* ingredient into confirmation. Hence I shall call it the *partly historical* or *logico-historical* (or *historical*, for short) *approach to confirmation*.

The main aim of this paper is to examine the acceptability of this logico-historical approach to confirmation. I shall begin with the paradox of confirmation, and with what I regard as the best solution to it, which was developed by Watkins using some of Popper's ideas. I will then distinguish three different ways of introducing an historical ingredient into confirmation, and assess their merits. I shall ask, in particular, whether once we move away from the logical approach, we can avoid introducing an undesirable subjectivity and arbitrariness into questions of evidential support—whether we can avoid saying that the evidential support of a theory depends upon personal or psychological facts about the man who proposed that theory.

2 'BACKGROUND KNOWLEDGE' AND THE PARADOX OF CONFIRMATION

To set the scene, let us restate the paradox of confirmation, and see how Watkins solves it by introducing the notion of 'background knowledge'. The paradox arises if we accept the following two principles:

---

[1] D. C. Stove [1960], p. 179; *cf.* Agassi [1961], p. 86, and Lakatos [1970], p. 123.

(*I*) Any positive instance of a hypothesis confirms that hypothesis[1];

(*E*) If e confirms H_1, and H_2 is logically equivalent to H_1, then e also confirms H_2.

As is well known, these two intuitively obvious principles lead to the intuitively absurd result that any object which is not a counter-instance to a universal hypothesis is a confirming instance to it. Principles (*I*) and (*E*) together render confirmation ridiculously *easy*. To use the usual example, any statement about the colour of any object will, provided it is not a report of a non-black raven, confirm 'All ravens are black'. As Goodman remarked, 'the prospect of being able to investigate ornithological theories without going out in the rain is so attractive that we know there must be a catch in it'.[2]

Solutions to the difficulty are legion, and I shall not attempt to canvas them all. Hempel's own solution to the paradox is simple—we must, he says, simply get used to it. According to Hempel, 'the impression of a paradoxical situation is not objectively founded; it is a psychological illusion'.[3] The illusion is fostered by two mistakes. First, we think that a statement like 'All ravens are black' is only about ravens, whereas in fact it asserts something about every object in the universe (that it is either not a raven or black). Second, we allow additional information to intrude. We think it odd that observations of objects inside a room can confirm 'All ravens are black' because we know in advance that there are no ravens inside rooms, and hence no non-black ones. But to take account of this additional knowledge is, says Hempel, to 'fail to observe the methodological fiction, characteristic of every case of confirmation, that we have no relevant evidence for *H* other than that included in *E*'. He says we must be 'careful to avoid this tacit reference to additional knowledge (which entirely changes the character of the problem)'.[4]

Watkins solves the paradox by recognising Hempel's 'methodological fiction' for what it is, namely, a fiction and one which begs the question in favour of the purely logical approach. Since the scientist is never in the position of having no information other than *E*, Watkins tries to take his additional information (his 'background knowledge') into account. And this does, as Hempel said it would, entirely change the character of the problem.[5]

[1] Whether e is a 'positive instance' of h is supposed to depend on the logical forms of e and h. Intuitively, an object which satisfies both the antecedent and consequent of a universal law will be described in a positive instance of that law. Hempel's own version of (*I*) reads: e confirms h if e logically implies the 'development' of h for the individuals mentioned (essentially) in e. For details see Hempel [1965], pp. 35–9.

[2] Goodman [1954], pp. 71-2.

[3] Hempel [1965], p. 18.

[4] Hempel [1965], pp. 19–20.

[5] Watkins' solution is contained in his [1964]. Popper and Watkins prefer to speak of 'corroboration' rather than 'confirmation' in order to distinguish their view from the purely logical approach. I shall not follow them in this.

Watkins begins with Popper's thesis that genuine confirmations can result only from the failure of genuine attempts to refute a hypothesis, from severe tests. But how are we to recognise a 'genuine attempt to refute' or a 'severe test'? Popper occasionally suggests that it is a psychological affair: for a test to be severe the experimenter who performs it must be sincerely trying to overthrow the theory tested. I have criticised this view elsewhere, and will say no more about it here except that the severity of tests must be an objective or impersonal matter, rather than a subjective or personal one.[1] Elsewhere Popper says the same, as when he claims that the severity of tests can be objectively compared once we take 'background knowledge' into account.[2]

According to this objective theory, the severest tests of a hypothesis are those which in the light of our background knowledge are most likely to refute it. And if background knowledge tells us that some possible observation or experiment will not refute our hypothesis, then making that observation or experiment will not be a severe test and so cannot result in a confirmation. Applied to the paradox of the ravens, suppose background knowledge tells us that ravens are not to be found in rooms.[3] Background knowledge therefore entails, about any particular object in a room, that it will not be a non-black raven—which is exactly what 'All ravens are black' predicts about that object. Thus background knowledge tells us that the prediction from our hypothesis is true, that we are not going to refute our hypothesis by observing objects in rooms. Since such an observation would not represent a severe test, it cannot result in a confirmation of 'All ravens are black'.

Suppose, on the other hand, that background knowledge tells us that the prediction from our hypothesis is likely to be false. To test such a prediction will be a severe test, and if the hypothesis passes the test it will be genuinely confirmed by the resulting evidence statement. The severest tests of all will be where background knowledge logically implies the negation of the prediction. A test of such a prediction will be a *crucial test* between the hypothesis and background knowledge.

Watkins makes it clear that if we assume that we have no background knowledge at all, then we are back in the Hempelian situation where all tests are equally severe (or rather, equally lacking in severity). Watkins also makes it clear that the introduction of (non-empty) background knowledge avoids the paradox by denying that every positive instance of a hypothesis confirms it. Finally, Watkins shows that *sometimes* so-called 'irrelevant instances' (evidence concerning non-ravens) may confirm 'All ravens are black'.[1] Anyone who finds it too counterintuitive to suppose that a positive instance need not confirm, while an irrelevant instance might, will have to reject Watkins' solution.

---

[1] See section 3 of my [1974].

[2] See, for example, Popper's [1963], p. 388.

[3] It almost goes without saying that our 'background knowledge' may in fact be false, and that taking it into account may lead us astray. But with nothing to guide him, the scientist would have no clue as to which of the infinitely many possible tests he should perform.

Watkins' solution also contains the principle that a hypothesis cannot be confirmed by the facts it was devised to explain. We need only assume that the facts a hypothesis is devised to explain are contained in the 'background knowledge' to that hypothesis. Clearly such facts cannot represent the results of severe tests, and so cannot confirm the hypothesis. It follows that a hypothesis can be confirmed only by testing *new independent predictions*. The intuitive idea that successful prediction is important is given a very sharp formulation: so far as evidential support is concerned, successful prediction is *all*-important.

This last point is contained in Popper's paper *The Aim of Science* published in 1957. Popper had earlier maintained that to be scientific a hypothesis must be empirically testable, it must exclude some observable state of affairs. But any hypothesis which explains the observed facts *e* is testable in this sense—for it excludes the observable state of affairs in which *e* is not the case. Thus Popper's testability requirement does not exclude explanatory systems which, though they yield known facts, do not lead to any new predictions. To avoid such explanations (*ad hoc* explanations, as he calls them), Popper strengthened his requirement of testability to a requirement of *independent testability*. Generations of scientists have been suspicious of *ad hoc* explanations. And Popper gives their suspicion a very strong rationale: only severe tests can confirm, only independent tests can be severe, so an *ad hoc* explanation (an explanation which is not independently testable) is unsatisfactory simply because it cannot genuinely be confirmed by empirical evidence.[2]

---

[1] Convincing examples of this are hard to find for trivial hypotheses like 'All ravens are black'. But suppose background knowledge tells us that all birds which have a certain property *P* are ravens. And suppose an observer spots a white bird with property *P*, call it *a*. Background knowledge together with *Pa* predicts that *a* is a raven—while 'All ravens are black' together with '*a* is not black' predicts that *a* is not a raven. Inspection of *a* to see if it is a raven will be a crucial test between the hypothesis and background knowledge, which might confirm the hypothesis. Yet if it does confirm it, the observation report will be of the form '*a* is a non-black non-raven'. Agassi has given a more plausible example of the same kind: *cf.* the account in Watkins [1964], pp. 108–9.

[2] It was Lakatos (in his [1968], pp. 376–8) who first emphasised the importance of the transition from testability to independent testability. As he says, it is independent testability which is important if our main interest is in the growth of knowledge.

Popper objected to *ad hoc* hypotheses already in his [1934], pp. 82–3,but there he used the term '*ad hoc*' in a more specialised sense. He pointed out that it is always possible to evade a refutation of a theoretical system by introducing auxiliary assumptions into that system. It would be unreasonable to exclude *all* such auxiliary assumptions, since some of them led to great theoretical triumphs (*e.g.* the discovery of Neptune). Therefore

Popper summarised these ideas in formulas, expressed in terms of logical probabilities, for the severity of tests and the degree of confirmation. The main idea is that the severity of a test resulting in evidence e on the hypothesis h in the light of background knowledge b is greater the more probable is e given h and b and the less probable is e given b alone: $S(e, h, b) = p(e, h \,\&\, b) - p(e, b)$. The degree of confirmation of h by e in the presence of b is defined so that it is proportional to the severity of the test. Thus if the severity of the test is zero, so is the degree of confirmation resulting from it.[1]

Now so far the notion of 'background knowledge' has been left rather vague. And as we will see, different variants of the historical approach to confirmation result from construing background knowledge in slightly different ways. But before proceeding, I should explain why I call a theory which takes into account background knowledge (however construed) a partly *historical* theory. I do so because once we have decided what *sort* of thing background knowledge is to contain, it will presumably be a historical task to determine its actual contents in a given case. Thus all variants of the historical approach will make the confirmation of a scientific theory somehow depend upon the historical setting in which that theory was proposed. Of course, once the actual content of background knowledge has been ascertained by historical investigation, the analysis of confirmation proceeds logically. But we investigate the logical relations between three things (theory, evidence, and background knowledge) and not two as in the purely logical approach.

How, then, is 'background knowledge' to be construed—what sort of thing does it contain? As soon as we pose this question, the spectre of subjectivism and relativism rises before us. For suppose we construe the

---

Popper formulates the following rule: 'The introduction of an auxiliary assumption into a system is permitted only if it does not reduce the degree of testability of the system as a whole—otherwise the auxiliary assumption is *ad hoc*'.

Now strictly speaking it is only *modifications* of a theoretical system, not auxiliary hypotheses themselves, which can be *ad hoc* in this sense. (The mere *addition* of another hypothesis to a system cannot remove a contradiction between that system and empirical evidence—hence the auxiliary hypothesis must *replace* one of the existing ones.) Clearly, Popper's old usage of the term '*ad hoc*', which refers to modifications of a system, is more restricted than his new one, which refers to systems as a whole. We might call them '*ad hoc*$_0$' and '*ad hoc*$_1$' to distinguish them. As we will see (*below*, note 2 to p. 20), Lakatos has introduced two further senses of the term.

[1] See Popper [1963], Appendix 2. Popper's formulas contain Mackie's 'inverse principle' or 'relevance criterion of confirmation'; *cf.* Mackie [1969]. Popper himself does not think his formulas can yield numerical degrees of confirmation. For them to do so, we would have to be able to assign numerical values to the function $p(a, b)$—and Popper does not think this can be done in all cases.

Moreover, these formulas do not do justice to the importance of crucial tests between a new and an old hypothesis. For background knowledge has to be pruned so that the conjunction '$h \,\&\, b$' is consistent, if the formulas are not to yield absurd results. For this reason, Lakatos modifies Popper's formulas: *cf.* Lakatos [1968], p. 375, note 2, and *below*, p. 17.

background knowledge to an hypothesis in a subjective or personal sense—as the collection of facts and theories which a particular scientist actually accepts as unproblematic while he considers that hypothesis. There is every reason to believe that different scientists will have different 'background knowledges' in this sense. And it will follow from the historical theory, thus construed, that the evidential support of a theory will depend upon which scientist we take to be considering that theory.[1]

The attempt to avoid this relativistic conclusion leads us to the first variant of the historical theory which I shall consider.

3 THE STRICTLY TEMPORAL VIEW OF BACKGROUND KNOWLEDGE

Suppose we insist that the 'background knowledge' to a given theory contains all the relevant experimental results, hypotheses, *etc.*, which are 'known to science' when that theory is proposed. The fact that different individuals will actually be aware of different portions of this background knowledge will cease to affect the confirmation of the new theory. Instead, we will have a ***strictly temporal*** view of background knowledge, and hence of independent testability. Facts 'known to science' before a hypothesis is proposed will not be able to confirm that hypothesis, since they will already be contained in background knowledge and cannot represent the results of severe tests. On this view, the only tests that can result in confirmations are tests of predictions of ***new effects***. And a theory is independently testable only if it predicts a ***novel fact***, a fact not 'known to science' when the theory was proposed.[2]

This view presupposes, of course, that the historian can settle the question of when a particular statement becomes 'known to science'. And it must be admitted that a certain penumbra of vagueness may surround this matter. Does a statement become 'known to science' when some scientist first thinks of it?—or when he first writes it down?—or when it is first published?—or perhaps when it is abstracted in a leading journal? Often it will not matter how we answer, for a very loose dating will suffice. And where it does matter, historians are quite used to settling such questions.[1] Given that we can make sense of the idea of a given body of knowledge in a certain field at a particular time, the strictly temporal view does enable us to avoid introducing extreme relativism into confirmation.

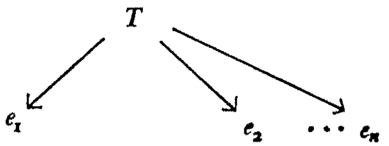
---

[1] Stove voiced essentially this objection to introducing background knowledge into confirmation in 1960: 'For consider two men, both of whom now defend a certain theory by appealing to certain observations, the theory and the observations being exactly the same in the two cases; but one of whom made the observations after, and because of, having conceived the theory, while the other conceived the theory after, and because of, having made the observations. Here, clearly, we must say, either that the corroboration, if any, which the observations afford is exactly the same in the two cases, or that the mental history of an investigator necessarily enters into the question of the evidential weight of his observations' ([1960], pp. 178–9).

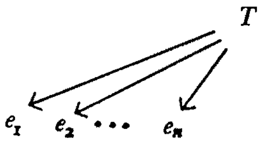
[2] According to Zahar ([1973], pp. 101–2), Lakatos was the first to adopt this strictly temporal view explicitly—and as we shall see, (*below*, p. 11), Lakatos certainly does try to soften some of the harsher consequences of such a view. Despite this, however, I have not been able to find where Lakatos explicitly adopts the strictly temporal view. Whenever he actually defines what he means by independent testability (or 'boldness', or the 'novelty of a fact') he adopts a different view, the one to be considered in section 5 *below* (I cite his definitions on p. 16, note 1).

But although an extreme relativity to persons is avoided, a less extreme though quite sharp historical relativity is introduced. Following Lakatos, I shall explain what this means by contrasting two schematic situations[2]:

Situation (1): Phenomenon e_1 is known, and theory T proposed which explains it. T is independently testable, since it predicts the unknown phenomena $e_2, \ldots, e_n$. These bold predictions are checked and confirmed. We have a marvellous success story.



Situation (2): The same phenomena $e_1, e_2, \ldots, e_n$ are all known in advance, and the same theory T proposed which explains them all. But now T is not independently testable, since it predicts no new phenomena. The historical theorist regretfully concludes that T is *ad hoc*.



---

[1] Zahar ([1973], p. 103, note 1) reports an amusing argument, due to John Worrall, which highlights the difficulties. Suppose scientist A devises theory T to account for the known facts e, and then uses T to predict the new fact f. A gets experimentalist B to check his prediction, and he confirms it. They send off their papers to different journals. But the theoretical journal takes longer to publish, so that the experimental confirmation of f appears in print before the theoretical prediction of it. On a strictly temporal view, can f confirm T?

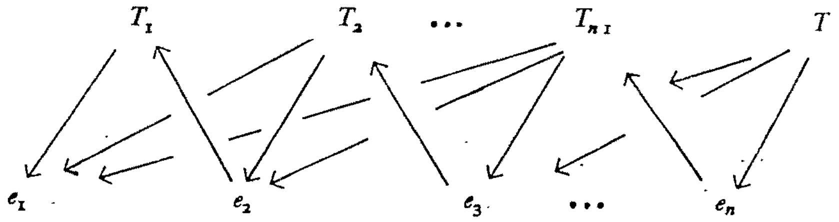
It can, of course. First of all, the two journals will probably print the dates of receipt of their articles, enabling T to get credit for having predicted f. Even if they do not (or if A's manuscript was delayed in the post), B may refer to A's theoretical prediction in his paper. Even if he does not, he will probably admit if asked that he performed his experiment as a result of hearing about A's prediction. And if he does not, A may be able to produce evidence (from manuscripts, letters, diaries) that his theory did predict f. In other words, the strictly temporal theorist must, in hard cases like this, adopt rather sophisticated historiographical methods.

[2] *Cp.* the schematic situations described by Lakatos in his [1970], pp. 151–2. Incidentally, in my diagrams time flows from left to right, downward arrows pointing left represent explanation, downward arrows pointing right represent successful prediction, and upward arrows represent refutation.

Now in both these situations we have exactly the same theory T and exactly the same body of evidence—the difference lies only in the order in which theory and evidence were discovered. For the logical theorist T will be equally well confirmed in both cases. But for the historical theorist T will be marvellously confirmed in the first situation, and cannot be confirmed at all in the second.[1]

It might be objected that my situation (*2*) never actually arises. For empirical facts are never simply accumulated as that situation supposes, but rather are only discovered by testing predictions from antecedently-held hypotheses. This interesting Popperian conjecture may well be true—I happen to think that, by and large, it is true.[2] But even if we accept it, we cannot avoid the historical relativity of confirmation. For we can replace situation (*2*) by:

Situation (*3*): Phenomenon e_1, and theory T_1 proposed which explains it. T_1 is independently testable and forbids phenomenon e_2. This new prediction is checked and found to be false. T_1 is rejected, and T_2 proposed which explains both e_1 and e_2. T_2 is also independently testable, and forbids e_3. Again the new prediction is checked and found to be false. The process continues until $e_1, \ldots, e_n$ are known, and T proposed which explains them all. But T is not independently testable, since it predicts no new phenomena. As in situation (*2*), the historical theorist regretfully concludes that T is *ad hoc*.



Here all the empirical facts (except for e_1) are discovered by testing predictions from antecedently held theories, precisely as the Popperian historical conjecture requires. Yet just as before, the historical theorist will have to say that T is marvellously confirmed in situation (*1*), and not confirmable at all in situation (*3*).[1]

---

[1] The inductivist regards theories which anticipate facts as premature speculation. So the inductivist teacher of science will present the history of theory T as if situation (*2*) were the case. But if he has Popperian students, he will only succeed in convincing them that T is *ad hoc*. If science actually proceeds, by and large, as in situation (*1*), then its inductivist virtues will be apparent only if it is taught unhistorically, while its Popperian virtues will be apparent only if it is taught historically.

Popper makes the interesting suggestion that the conventionalist view of physical theory arises from taking an unhistorical view of its development—see his [1963], p. 240.

[2] But *cp. below*, note 1 to page 18.

Those whose intuitions favour a purely logical approach to confirmation will, no doubt, find this kind of historical relativity absurd. But the purely logical approach has some absurd results of its own. It seems that in this area *some* of our intuitions have to go—and I for one am prepared to accept that the historical background to a theory might make a difference to its confirmation.

But ought strictly *temporal* considerations count for so much as this variant of the historical theory suggests? The strictly temporal view of independent testability is certainly a very stringent one. It means that Galileo's and Kepler's laws (or facts about the tides, or the precession of the equinoxes) cannot confirm Newton's theory, simply because they were 'known to science' before that theory was proposed.[2] It means that the Michelson–Morley experiment cannot confirm the Special Theory of Relativity, nor can the anomalous precession of Mercury's perihelion confirm the General Theory. And since they too were known in advance, Balmer's empirical formulas for the emission spectrum of excited hydrogen cannot confirm Bohr's theory of the hydrogen atom.

In connection with this last example, Lakatos tries to water down the strictly temporal view of background knowledge. He suggests that if a known fact is 'reinterpreted' in the light of a new theory, it can turn into a novel fact for that theory: 'we should certainly regard a newly interpreted fact as a new fact, ignoring the insolent priority claims of amateur fact collectors'.[3]

Now it will hardly do, especially from a methodologist who professes to take the judgements of scientists seriously, to consign a great experimentalist like Balmer to the flames as an 'amateur fact collector'. Moreover, as Zahar points out, Lakatos's watering-down of the notion of a novel fact actually obliterates it altogether.[4] Any deduction of an old fact from a new theory can be said to involve a 'reinterpretation' of that fact. And now any *ad hoc* theory can claim evidential support from the old, but 'newly interpreted', facts which it explains.

---

[1] In terms of Popper's three requirements for the growth of knowledge (to be discussed in section 6), in situation (*3*) each of the T_i $(1 \leqslant i \leqslant n-1)$ satisfies the second requirement (independent testability) but fails the third (independent confirmation), while T fails to surmount even the second hurdle. Yet so far as the evidence goes, T may be true while all its predecessors are false!

[2] In fact, of course, Newton predicted deviations from Galileo's and Kepler's laws, most of which represented novel facts. Since Newton's theory was devised to explain Galileo's and Kepler's laws, it is the paradigm of a non-*ad hoc* theory, since it corrects its own *explanandum*. See Popper's [1972], p. 202.

[3] Lakatos [1970], p. 157; the whole sentence is italicised in the original.

[4] See Zahar [1973], p. 102. Noretta Koertge generously chose to ignore what she called Lakatos's 'lapse into the we-live-in-a-different-world-after-a-revolution syndrome', and insisted that if he 'wants to count a research-programme as progressive when it is found to predict a known result, he should just say so'; see her [1971], p. 171, note 5.

Zahar rightly deplores this wholesale retreat—yet he still wants the Michelson–Morley experiment to confirm the theory of relativity. He therefore proposes a new and ingenious view of independent testability or of the novelty of a prediction. This brings me to the second variant of the historical theory.

4 THE HEURISTIC VIEW OF BACKGROUND KNOWLEDGE

Zahar suggests that the novelty of a fact should not depend merely on temporal considerations. Instead, an old fact can be novel with respect to a new theory *provided that theory was not devised to explain it.* Hence temporal novelty is a sufficient condition for novelty in Zahar's sense, but not a necessary one. If the Michelson–Morley experiment played no role in the construction of the Special theory, then it can still confirm that theory (and similarly for the other examples mentioned above). Or, to revert to our schematic situations, T is no longer automatically *ad hoc* in situations (2) and (3). It all depends upon whether the scientist who devised T made use of all the known facts $e, \ldots, e_n$ in doing so—if he did T is *ad hoc*, if he did not it is independently testable.

Zahar's proposal involves a *heuristic view of background knowledge.* The background knowledge to a new theory does not contain all the previously known facts which it explains, but rather only the previously known facts which it was designed to explain and which played a heuristic role in its construction.

Now Zahar is quite right to insist that the facts which a new theory is *devised* to explain are likely to be less extensive than the previously known facts which it does actually explain. There is even a general argument why this is so: that one or more apparently unrelated facts are connected with each other often becomes clear only *in the light of* some theory—and so it is unlikely that that theory was devised to account for all of them. Thus we must distinguish the actual *explanandum* for a theory from the previously known facts which it explains.

But the novelty of Zahar's methodological proposal is that this distinction should have repercussions for the new theory's evidential support. According to Zahar, the way in which a scientist happens to have arrived at his theory can affect the extent to which that theory is confirmed by empirical evidence. To assess the evidential support of a theory 'one has to take into account the way [it] is built and the problems it was designed to solve'.[1]

---

[1] Zahar [1973], p. 103 (italicised in the original).

Thus confirmation depends, not on comparatively public historical matters like the dates of discoveries, but on private biographical details about the man who proposed the theory in question. Zahar is well aware that his view 'implies that the traditional methods of historical research are even more vital for evaluating experimental support than [anybody] had previously suggested. The historian has to read the private correspondence of the scientist whose ideas he is studying . . . to disentangle the heuristic reasoning which the latter used in order to arrive at a new theory'.[1] According to Zahar, then, the private correspondence, diaries, manuscripts, *etc.*, of an individual scientist are no longer merely of historical or heuristic interest—their contents may also help to determine the evidential support of his theory.

Now the first thing to be said about this proposal is that it will make the evidential support of a theory rather *difficult to determine*. Scientists are notoriously reticent about the routes to their discoveries. Most of them assume that the way in which a theory is arrived at is irrelevant to its value, including its evidential support. Zahar disputes this assumption (and the distinction between the 'context of discovery' and the 'context of justification' which philosophers of science have erected upon it). A scientist's own account of his discovery, often written when he is too old to do anything better than reminisce about 'how I did it', will often be distorted by philosophical preconceptions (including, perhaps, Zahar's should it gain currency amongst scientists). But Zahar can, of course, reply to this that the historian is used to dealing with such difficulties.

A more important question is whether, by departing from the austere standards of the strictly temporal view, Zahar can avoid making the evidential support of a theory a person-relative affair. Suppose two scientists, A and B, independently and at about the same time, propose the hypothesis h. And suppose that A devised h to account for the known facts e_1 and e_2—while B, less *au fait* with current literature, devised h only to account for e_1. Does Zahar's view entail that e_2 confirms h as proposed by B, but does not confirm h as proposed by A? The historian might salute B's ingenuity (or, perhaps, remark upon his good luck). But will Zahar say that B's ignorance of the literature means that h as proposed by him has more empirical support than the same hypothesis as proposed by A?

I suspect that Zahar would wish to avoid conclusions like this one. For he repeatedly insists that disentangling the heuristic route to a discovery is not, as he calls it, a 'psychologistic' affair.[2] Thus, he might say, a scientist can be *aware of* some extant proposition without it *playing a heuristic role* in his theorizing (can he, I wonder, be unaware of a proposition which does play a heuristic role?). For example, even if Einstein was aware of the results of the Michelson–Morley experiments, this does not mean that they played any role in the construction of his theories.

---

[1] Zahar [1973], pp. 103–4.

[2] Thus, for example, Zahar insists that our purpose in reading the correspondence of the scientist who invented the theory we are interested in 'will not be to delve into the psyche of the scientist, but to disentangle the heuristic reasoning which the latter used in order

But how successful are these disclaimers? Is disentangling a process of heuristic reasoning really so 'non-psychologistic' as Zahar wants it to be? Zahar provides us with a fascinating example—the heuristic route to Einstein's theories of relativity. Now I shall not dispute the accuracy of Zahar's account, for I am hardly competent to do so. But I do claim that, unless it is a mere fairy-story with no pretensions to historical accuracy, it must contain historical assertions about the actual beliefs, principles used, *etc.*, of a particular person.

For example, according to Zahar two metaphysical principles and two empirical facts played an important heuristic role in the genesis of the theory of relativity. The metaphysical principles are: Nature is simple or coherent (its associated methodological prescription says that adequate theories should be simple or coherent); and there are no accidents in Nature (its associated methodological prescription is that observed symmetries must be matched by theoretical symmetries). The empirical facts are the results of electromagnetic induction experiments, and the equality of inertial and gravitational mass.[1] Now Zahar must claim, it seems to me, that Einstein actually made use of the metaphysical principles, and actually considered the empirical facts. And such claims are psychological or personal claims about a particular historical figure. But if the evidential support of a theory depends upon the heuristic route to it, and the heuristic route to it involves facts about the man who took that route, I cannot see how Zahar can avoid the conclusion that the evidential support of a theory depends upon facts about the man who proposed it. If different scientists take different heuristic routes to the same theory, then the evidential support of that theory as proposed by one of them might be different from its evidential support as proposed by the other. In short, Zahar's view makes confirmation a person-relative affair.

---

to arrive at a new theory' ([1973], p. 104). Later, when he comes to his case-study, Zahar begins by asking why Einstein objected to classical physics—and he takes pains to reassure us that 'the answer to this question will not be a psychologistic answer; I shall not for example be indulging in speculations about Einstein's childhood' ([1973], p. 223).

[1] See Zahar [1973], pp. 223–7. Incidentally, Einstein's heuristic (as reconstructed by Zahar) bears little resemblance to the 'positive and negative heuristic' of Lakatos's methodology of scientific research programmes. Einstein's heuristics are closer to the sort of 'influential metaphysics' from which Lakatos disassociates himself (*cf.* his [1970], pp. 183–4). This obviously bears upon Zahar's claim that he is presenting a case-study of the 'methodology of scientific research programmes'.

At this point Zahar might be tempted to invoke Lakatos's notion of ***rational reconstruction***. He might say that he is not interested in whether Einstein actually considered the empirical facts he mentions in the light of the metaphysical principles he formulates. He is only 'rationally reconstructing' a heuristic which Einstein might have, or perhaps ought to have, followed. Kuhn has already objected that Lakatos's rational reconstructions are not so much history as philosophy fabricating examples.[1] More importantly, since philosopher-historians of science can provide different rational reconstructions of the genesis of the same theory, the evidential support of a theory will now depend upon which philosopher-historian of science we consider. I doubt that scientists themselves will be prepared to hand over to philosophers of science the task of assessing the evidential support of their theories.

This is not to deny that the notion of 'rational reconstruction' has a part to play in the history of science. As Mark Twain said, in history nothing ever happens in the right place or at the right time—and it is the job of the historian to remedy this defect. The historian can, and does, rationally reconstruct an episode and then compares his reconstruction with the actual history—and the exercise can be most illuminating. But when a historian of science rationally reconstructs the problem-situation for a new theory, surely the most sensible reconstruction will consist of the known problems actually solved by the theory in question. And this will lead us back to a strictly temporal view of 'novelty' or of independent testability.

But if the strictly temporal view is too austere, and Zahar's heuristic view too relativistic, can we preserve any of the insights of the historical approach to confirmation? I think that we can—which brings me to the third variant of the historical approach.

5 BACKGROUND KNOWLEDGE OR A BACKGROUND THEORY?

This third variant starts from the basic idea that scientists operate with competing theories, and that they use empirical evidence to try to decide between them. The basic question is not so much 'Does evidence e confirm hypothesis h?' but rather 'Does evidence e support h_1 more than h_2?'. This suggests that in assessing the evidential support of a new theory we should compare it, not with 'background knowledge' in general, but with the old theory which it challenges. Lakatos, who developed this third variant, calls the old theory the 'background theory' or the 'touchstone theory'.[2]

According to this view, a new theory is independently testable (or predicts a 'novel fact') if it predicts something which is not also predicted by its background theory.[1] Hence there are two kinds of independent or novel predictions, tests of which are severe and might result in confirmations. First, there are predictions which *conflict* with the predictions of the background theory—tests of these will be crucial tests between the new theory and the old.[2] Second, there are predictions concerning phenomena about which the background theory predicts nothing at all—tests of these will also be independent tests and severe ones (in the light of 'background knowledge' construed as a competing theory).

---

[1] Kuhn [1971], p. 143.

[2] Lakatos [1968], pp. 375–90.

This third variant gives different results than the other two variants. First, it only partially endorses the view that a theory cannot be confirmed by facts contained in its *explanandum*. For suppose that the (rationally reconstructed) *explanandum* for a new theory contains the successes and failures of the old theory. The 'failures' being known empirical facts which contradict or cannot be explained by the old theory. According to our third variant, such facts can confirm the new theory despite being known before that theory was proposed. Facts which contradict the old theory and are predicted by the new represent the results of crucial tests between them—and so, it seems, a crucial test between T_1 and T_2 can be performed even before T_2 is proposed![3]

So we see that this third variant does not involve a strictly temporal view of the 'novelty' of a prediction. A known fact can confirm a new theory if it contradicts, or cannot be explained by, its predecessor. And a fact discovered by testing a prediction from the new theory will not confirm it if it can also be predicted by the old theory. Thus temporal novelty of a factual discovery is neither sufficient for it to confirm a theory, nor is it necessary.

Our third variant might agree with some of Zahar's conclusions, but it gives different reasons for them. The Michelson–Morley experiment can confirm the Special Theory of Relativity, and the Mercury perihelion the General Theory, because they refuted the previous theories. That these experimental results played no role in the construction of Relativity Theory may be of great *heuristic* interest—but according to our third variant it is quite irrelevant to questions of evidential support.

---

[1] Or as Lakatos put it, the new theory should have potential falsifiers which are not also potential falsifiers of the old theory ([1968], pp. 376–7). The same view is expressed differently in his [1970], p. 116: to be independently testable or have 'excess empirical content' over its predecessor a new theory must predict '*novel* facts, that is, facts improbable in the light of, or even forbidden by' the old theory. Clearly, a 'novel fact' thus defined may very well be a *known* fact—and hence Lakatos does not *explicitly* adopt a strictly temporal view of novelty (but see the next footnote).

[2] At one point Lakatos says that where the old theory T_1 has been refuted by evidence e we should make the conjunction T_1 & e the background theory for any new theory. This proposal would take him close to a strictly temporal view of independent testability. But the proposal is unsatisfactory. It means that crucial tests between the old theory and the new need not be severe ones. Indeed, since T_1 & e is, by assumption, inconsistent, and since inconsistent premises imply anything, the proposal means that with such a 'background theory' *no* test can be severe. The proposal was, therefore, a slip and I shall say no more about it (*cf.* Lakatos [1968], p. 375, last paragraph).

[3] This should cause no surprise. Of course, *before* T_2 is proposed such an experiment will not *appear* to be crucial. As Lakatos (and others) have emphasised, before the new theory is available refutations of the old are classed as 'unexplained anomalies'—their crucial character only becomes apparent with hindsight (see, for example, Lakatos [1970], pp. 154–9).

To see that the third variant will give different results from Zahar's, imagine the following situation. The existing theory T_1 explains e_1 and e_2 but cannot explain e_3. A new theory T_2 is proposed and e_1 and e_3 play a heuristic role in the construction of it, while e_2 does not. Now according to the third variant e_3 will confirm T_2, while according to Zahar it will not. And according to the third variant e_2 will not confirm T_2, while according to Zahar it will. (On a purely historical view, none of e_1, e_2, e_3 can confirm T_2.) As we can see from this, Zahar's view threatens to make the confirmation of a new theory rather easy: for if none of the known facts which it entails played a role in its construction, then they will all confirm it irrespective of whether they also confirm existing theories. Yet scientists themselves regard genuine confirmations of, say, Relativity Theory as very difficult to come by indeed—which suggests that by genuine confirming evidence for a new theory they mean evidence which is not also explained by existing theories.

Lakatos uses the idea that background knowledge should be replaced by a background theory to simplify and improve Popper's definition of the severity of tests. The severity of a test resulting in e on the hypothesis h in the light of the 'touchstone theory' t can be given by:

$$S(e, h, t) = p(e, h) - p(e, t).^1$$

As this formula makes clear, according to the third variant confirmation depends solely upon the logical relations between the evidence and the competing theories in question. Facts about the man who proposed the new theory, or about the heuristic route he followed, will be quite irrelevant.

In view of this, one may wonder whether the third view is really entitled to be called a variant of the historical approach at all. And it certainly is possible to use it as a theory of the *selective confirmation* provided by the evidence e for the theories T_1 and T_2. Used in this way, the background theory T_2 to a given theory T_1 can be chosen quite arbitrarily.

---

[1] Lakatos [1968], p. 382. Where both h and t predict e, $S(e, h, t)$ is 0. Where h predicts e but t does not, $S(e, h, t)$ is $1-r$ where r is $p(e, t)$. Where h predicts e but t predicts not-e, so that $p(e, t)$ is 0, $S(e, h, t)$ is 1. Thus Lakatos's definition gives maximum severity to crucial experiments between h and t, while Popper's places no premium on crucial experiments at all (see *above*, note 1 to page 7).

One small anomaly can be easily repaired. If h predicts e and t predicts not-e, and if we let the evidence statement of the formula be not-e, it yields a severity of *minus* 1. So we should take the absolute value of $p(e, h)-p(e, t)$ to give us severity.

But our third view can also be used as a variant of the historical theory, provided we require that the 'background theory' for a new theory must be the best available theory actually present in the field. Then it will be a historical task to locate that theory in any given case. And if it is used in this way, our third variant, like the first, introduces a *historical relativity* into confirmation.

To see this, let us return to our three schematic situations. In situation (*1*), T will be confirmed by $e_2, \ldots, e_n$ since it predicts them and (by assumption) no existing theory does the same. But in situation (*2*), evidence $e_2, \ldots, e_{n-1}$ will not confirm T because all of it is also predicted by the background theory T_{n-1}. In situation (*3*), only e_n confirms T—because it refutes its background theory. Thus in situation (*3*) T will not be confirmed by evidence which might have confirmed it had the historical situation been different, and T not been preceded by T_{n-1}. For example, according to this third variant, because Einstein had the misfortune to be preceded by Newton, his theory cannot be confirmed by all the evidence which it predicts, but which is also predicted by Newton's theory.

Our schematic situations do pose one minor problem. Does e_1 confirm T in situation (*1*)—and do $e_1, \ldots, e_n$ confirm T in situation (*2*)? In other words, what are we to regard as the 'background theory' to the *first* testable theory in a given field? The natural answer, it seems to me, is to assume that in such a case the background theory is the empty or tautologous theory. If we do so, then the first testable theory in a given field will be confirmed by all the phenomena which it explains (whether or not they were known in advance, and whether or not they played a heuristic role in the construction of the theory). The Bohr–Balmer case *may* be a case in point: if Bohr's theory was the first explanation of Balmer's empirical formulas, then it can be confirmed by them.[1] We see, then, that the first theorist in a given field has a marked advantage as far as evidential support is concerned—while the better supported are existing theories, the more difficult it is to improve upon them. This is a point to which I shall return.

To sum up our discussion of the three variants of the historical theory. The first variant takes 'background knowledge' to include everything 'known to science' prior to the proposal of the new theory—it has the counterintuitive result that a new theory cannot be confirmed by known facts which refute its predecessor. The second variant takes 'background knowledge' to include only known things which played a heuristic role in the construction of the new theory—it has the counterintuitive result that how well a theory is confirmed might depend upon who we take to have proposed it. The third variant takes 'background knowledge' to include only the best existing competing theory—and for my money it is the best version of the historical approach to confirmation.

---

[1] In this way we arrive at the same result as Lakatos (see *above*, p. 11) but for a very different reason. Incidentally, Balmer's spectroscopic investigations seem to refute the Popperian historical conjecture I mentioned earlier (see *above*, p. 10). For they were not undertaken as *tests* of any prior theory. Balmer did, of course, make use of all sorts of theories (for example, those involved in the construction of the spectroscope)—but to make use of a theory and to test it are not the same.

Whichever variant we prefer, it will have the result that a theory need not be empirically confirmed by evidence which shows that one of its predictions is true. In case anybody still cannot stomach this consequence, I shall conclude by trying to make it a little more palatable.

6 REQUIREMENTS FOR THE GROWTH OF KNOWLEDGE

To do this, we need only remember that empirical confirmation is just *one* of the requirements for a good theory. Popper, for example, lists three requirements that a new theory must satisfy if it is to represent progress in science. The first requirement is that the theory be *simple* or *unified*. The second is that it be *independently testable* or *not ad hoc*. And the third is that it be (independently) *confirmed*—it 'should pass some new and severe tests'.[1]

Now we have seen that the notions of an independent test and an independent confirmation can be interpreted in (at least) three different ways—and it is not clear from Popper's formulations which he intends.[2]

Moreover, Popper's third requirement has occasioned some controversy. He actually requires, not merely that the new theory be independently confirmed, but that it pass the *first* independent tests to which it is subjected. I agree with Lakatos that this was careless of him.[3] Furthermore, Popper claims, rather mysteriously, that his third requirement is necessary to eliminate *ad hoc* theories:

> '... it is clear that the mere fact that the theory is ... independently testable cannot as such ensure that it is not *ad hoc*. This becomes clear if we consider that it is always possible, by a trivial stratagem, to make an *ad hoc* theory independently testable, *if we do not also require that it should pass the independent tests in question*: we merely have to connect it (conjunctively) ... with any testable but not yet

---

[1] Popper [1963], pp. 240–8. If we construe 'independently testable' in accordance with the third variant, then Lakatos's 'acceptability$_1$' and 'acceptability$_2$' (*cf.* his [1968], pp. 375–90) correspond to Popper's second and third requirements respectively.

[2] For example, Popper's formulation of his second requirement is ambiguous between the first two variants of the historical theory: 'For, secondly, we require that a new theory should be *independently* testable. That is to say, apart from explaining all the *explicanda* which the new theory was designed to explain, it must have new and testable consequences (preferably consequences of a *new kind*); it must lead to the prediction of phenomena which have not so far been observed' ([1963], p. 241). Later Popper hints also at the third variant: see, for example, his [1963], pp. 246–7.

[3] *Cf.* Lakatos [1968], p. 388.

> tested fantastic *ad hoc* prediction which may occur to us (or to some science fiction writer).
>
> Thus our third requirement, like the second, is needed in order to eliminate trivial and other *ad hoc* theories.[1]

This is a bad argument for a good principle. First of all, the term '*ad hoc*' is used in two very different senses (at least—testable predictions are also described as '*ad hoc*'). First, a theory is *ad hoc* if it is not independently testable. Second, an independently testable theory is *ad hoc* if it is not independently confirmed.[2]

But more important than the obscurity of the argument is the fact that Popper's third requirement cannot do the job Popper here gives it, nor is it necessary for it to do so. For suppose we do conjoin a theory T which is not independently testable with an arbitrary testable prediction P. And suppose that the prediction gets confirmed. Then the 'theory' T & P will *satisfy* Popper's third requirement. I think, however, that scientists would regard this as an unsatisfactory 'theory', and would not even bother to submit it to test, And they would do so because it does not satisfy the requirement that Popper places first in his list, the requirement of simplicity. Popper formulates this as follows:

> The new theory should proceed from some *simple, new, and powerful unifying idea* about some connection or relation ... between hitherto unconnected things ... or facts ... or 'theoretical entities' ... This *requirement of simplicity* is a bit vague, and it seems difficult to formulate it very clearly ... [It is] intuitively connected with the idea of a unified or coherent system or a theory that springs from *one* intuitive picture of the facts ...[3]

Now by assumption the arbitrary conjunction T & P is not simple or unified or coherent, and will already be excluded by Popper's first requirement. We do not need his third requirement to do the job, nor can it really do so, as we have seen.

Are we then to dispense with Popper's third requirement, since the argument he gives for it is a bad one? I do not think so, for Popper gives another very simple argument in its favour. Scientists search for truth, and want a new theory to have more truth in it than its predecessor. And a new theory will not have more (empirically checkable) truth in it than its predecessor unless at least one of its independent predictions gets confirmed. Hence the requirement of independent confirmation.[1]

---

[1] Popper [1963], p. 244.

[2] Lakatos labels these two senses of the term '*ad hoc*$_1$' and '*ad hoc*$_2$' respectively (in his [1968], p. 389, note 1). I shall not follow him in this; I see no point in calling a theory *ad hoc*$_2$ when what you mean is that it has not been independently confirmed. Alas, having introduced the subscripts, Lakatos warms to the device and makes the terminological thicket even thicker. He introduces the term '*ad hoc*$_3$' to describe theories (or rather, theoretical developments) which are not simple or unified (*cf.* his [1970], p. 175). Zahar goes even further, and suggests, in effect, that we add a further temporal subscript to Lakatos's '*ad hoc*$_2$' (see his [1973], p. 101, note 1).

[3] Popper [1963], p. 241 (I have amalgamated the text and a footnote to it).

Now I promised to use these three requirements to try to make the historical approach to confirmation a little more palatable. It is very easy to do this. We have *three* requirements for a good theory—and ideally, of course, a new theory will satisfy all three of them. But the requirements are exacting ones, and the more successful science has been in fulfilling them to date the more difficult it becomes to fulfil them in the future. The deeper and more comprehensive our present theories, the more difficult it is genuinely to unify them into a new theory which is not only independently testable, but which passes some of these tests.[2] I suggest, therefore, that failure to satisfy *all* these requirements should not be regarded as a condemnation of a theory. This is pretty trite. A theory, independently testable or not, which *genuinely* unifies previously disjoint facts or theories can be a theoretical triumph, because of the theoretical problems it solves and the theoretical suggestions it makes. An independently testable theory which never gets confirmed can still contribute by leading to new factual discoveries and posing new problems for the theoretician. Of course, the *best* theories will satisfy all three of the requirements.[3]

At any rate, I do not think that historical theorists can have it both ways. By allowing only independent tests (however construed) to confirm, they make confirmation more difficult. And it becomes more and more difficult the more successfully science has met their requirements in the past. So they should not be too hard on theories which, by their austere standards, are *ad hoc*.[4] Of course, as moralists they can continue to preach *all* the scientific virtues—but like moralists they should not be too swift to condemn that which has only some of them.[1]

---

[1] For this second argument see Popper's [1963], pp. 245–6. In stating it I made two large assumptions. First, that the scientist's decisions about the success or failure of predictions are correct. Second, that the new theory does contain all the true testable consequences of the old.

I have also avoided the very real difficulties in giving an adequate definition of Popper's notion of *verisimilitude*. As Tichý has shown, if the new theory and the old are both false, then on Popper's own definitions the one cannot have greater verisimilitude than the other—see Tichý [1974].

[2] Newton is supposed to have remarked: 'If I have seen further it is by standing on the shoulders of giants'. Usually this is seen as an expression of his modesty—sometimes of his *false* modesty. I suspect that it may have been a straightforwardly megalomaniac remark: that it is more difficult to see further than a giant than it is to see further than a dwarf, and that it was very clever of him to have done so.

[3] I do not think Popper would disagree with all this. He says that an *ad hoc* theory which represents a genuine unification may be a very great achievement ([1963], p. 241). And he says that a theory which satisfies his first two requirements but not the third can be 'an important contribution to science' ([1963], p. 243; his example is the Bohr–Kramers–Slater theory).

[4] Of course, there is always a *suspicion* that an *ad hoc* theory has been merely 'cooked up' in some trivial way to yield existing results. But an independently testable theory may also have been 'cooked up'—and so may an independently confirmed one (see *above*, p. 20). Philosophers of science have made little progress in explaining how a 'cooked up' theory

7 CONCLUSION

If I were asked to sum up, I would say this. The logical approach to confirmation renders it so easy that it ceases to be important. While the historical approach renders it so difficult that it should cease to be all-important.

University of Otago,
Dunedin, New Zealand

REFERENCES

AGASSI, J. [1961]: 'The role of corroboration in Popper's methodology', *Australasian Journal of Philosophy*, 39, pp. 82–91.

BLAKE, R. M. [1960]: 'Theory of Hypothesis among Renaissance Astronomers', in E. H. Madden (ed.): *Theories of Scientific Method: The Renaissance through the Nineteenth Century*, pp. 22–49.

DUHEM, P. [1954]: *The Aim and Structure of Physical Theory.*

FEYERABEND, P. [1970]: 'Problems of Empiricism, II', in R. G. Colodny (*ed.*): *The Nature and Function of Scientific Theories*, pp. 275–353.

GOODMAN, N. [1954]: *Fact, Fiction and Forecast.*

---

differs from a coherent or unified one (though not much sophistication is needed to recognise that the results of the philosopher's tacking arguments are 'cooked up'). But working scientists do have means of distinguishing them.

[1] The fact that there is more than one *desideratum* for a good theory (and hence more than one associated methodological rule) has some interesting repercussions. For in a given case the *desiderata* or rules may give conflicting results: T_2 may be more coherent than T_1, while T_1 is better supported by evidence than T_2. In such a situation two scientists might make their choices in different ways. And (short of attempting to *weight* his *desiderata*) the methodologist will have to say that both of them are acting rationally. Thomas Kuhn has been making the point for years—it is actually all that his famous 'incommensurability thesis' really amounts to (see my [1971], p. 294).

I am also tempted to try to bring Feyerabend back from beyond the irrationalist pale. He claims to be an *epistemological anarchist*. He says that, for any methodological rule yet proposed, he will find a scientist who violated it—and a good thing too. And he concludes that the only sensible rule is: there are no rules, anything goes.

But if there are multiple desiderata (multiple rules), and they give conflicting results, an individual may quite rationally set one aside in favour of another. Inspection of Feyerabend's case-study of Galileo (in his [1970]) will, I think, reveal that this is exactly what happened: Galileo was no 'anarchist' but his rules gave him conflicting results and he had to back some and violate the others temporarily. Thus Feyerabend's premise (each rule has a fruitful historic violation) does not, even if correct, establish his conclusion (there should be no rules). Similarly one cannot establish moral anarchism by showing that, for any moral rule, there are situations where it ought to be violated—what this would show is only that no single rule is so sacrosanct that it overrides all others in all situations.

Feyerabend's 'epistemological anarchism' has taken him so far to the left of the other Popperians that he is in danger of falling off their platform altogether. The above will rescue him from this sad fate. He will, no doubt, be furious at the rescue act, and may make even more frantic efforts to fall off. Which is why I try to rescue him, for his cavortings are always amusing and sometimes instructive.

HEMPEL, C. G. [1965]: *Aspects of Scientific Explanation and other essays in the Philosophy of Science.*

KEYNES, J. M. [1921]: *A Treatise on Probability.*

KOERTGE, N. [1971]: 'Inter-theoretic criticism and the Growth of Science', in R. Buck and R. S. Cohen (*eds.*): *Boston Studies in the Philosophy of Science*, **8**, pp. 160–73.

KUHN, T. S. [1971]: 'Notes on Lakatos', in R. Buck and R. S. Cohen (*eds.*): *Boston Studies in the Philosophy of Science*, **8**, pp. 137–46.

LAKATOS, I. [1968]: 'Changes in the Problem of Inductive Logic', in I. Lakatos (*ed.*): *The Problem of Inductive Logic*, pp. 315–417.

LAKATOS, I. [1970]: 'Falsification and the Methodology of Scientific Research Programmes', in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge*, pp. 91–195.

MACKIE, J. L. [1969]: 'The Relevance Criterion of Confirmation', *British Journal for the Philosophy of Science*, **20**, pp. 27–40.

MEDAWAR, P. [1967]: *The Art of the Soluble.*

MILL, J. S. [1843]: *A System of Logic.*

MUSGRAVE, A. E. [1971]: 'Kuhn's Second Thoughts', *British Journal for the Philosophy of Science*, **22**, pp. 287–97.

MUSGRAVE, A. E. [1974]: 'The Objectivism of Popper's Epistemology', in P. A. Schilpp (*ed.*): *The Philosophy of K. R. Popper, forthcoming.*

POPPER, K. R. [1934]: *Logik der Forschung* (English translation, 1959).

POPPER, K. R. [1963]: *Conjectures and Refutations.*

POPPER, K. R. [1972]: *Objective Knowledge.*

STOVE, D. C. [1960]: 'Review of *The Logic of Scientific Discovery* by K. R. Popper', *Australasian Journal of Philosophy*, **38**, pp. 173–87.

TICHÝ, P. [1974]: 'On Popper's Definitions of Verisimilitude', *British Journal for the Philosophy of Science*, **25**.

WATKINS, J. W. N. [1964]: 'Confirmation, the Paradoxes, and Positivism', in M. Bunge (*ed.*): *The Critical Approach to Science and Philosophy*, pp. 92–115.

WHEWELL, W. [1840]: *Philosophy of the Inductive Sciences.*

ZAHAR, E. G. [1973]: 'Why did Einstein's Programme supersede Lorentz's?', *British Journal for the Philosophy of Science*, **24**, pp. 95–123 and 223–62.

Brit. J. Phil. Sci. 25 (1974), 25–82 *Printed in Great Britain*

Discussions

ZAHAR ON EINSTEIN

1 Elie Zahar applies the methodology of research programmes to the transition from Lorentz to Einstein.[1] Using a new and highly imaginative concept of novelty he shows (*i*) that Lorentz's programme progressed up to 1905; (*ii*) that it was rational for Lorentz to pursue it after publication of Einstein's paper; (*iii*) that the programme was superseded by relativity in 1915, with the explanation of the precession of the perihelion of Mercury; (*iv*) that those who joined Einstein after 1905 and before 1915 had objective reasons for doing so. He also makes other observations, some of them very interesting.

Zahar's account supersedes the existent accounts, it is more sophisticated than any one of them, both logically and historically. Indeed, we learn from him, as we learn from Lakatos, that history cannot be done without a considerable amount of conceptual finesse. Even the problems to which Zahar's account gives rise are more colourful than the dreary problems his predecessors have to face. Let me mention some of them.

2 The methodology of research programmes develops standards for the evaluation of (scientific, or, more generally, conceptual) *change*. The standards apply to research programmes, not to individual theories; they judge the evolution of a programme over a period of time, not its shape at a particular time; and they judge this evolution in comparison with the evolution of rivals, not by itself. A research programme is called 'progressive', if it makes predictions that are confirmed by subsequent research and thus lead to the discovery of novel facts. It is called 'stagnating' if it makes no such predictions, but is reduced to absorbing material that was discovered with the help of its rivals. Thus the first step in the application of the methodology of research programmes is to present the research programmes which are to be compared.

3 Zahar chooses as the hard core L' of Lorentz's programme, Maxwell's equations (*in vacuo*), Newton's laws of motion together with the Galilean transformations, and the Lorentz force. He should have added atomism, for Lorentz designed his research programme (which I shall call L) for the purpose of explaining electromagnetic phenomena *in matter*. Formalists may abstract from such elements when discussing physical theories, followers of the methodology of research programmes must not.

Zahar never spells out the hard core of Einstein's research programme. One may guess, however, that it consists of some principles which are shared by the special and the general theories of relativity. I shall call these principles R. This again omits important elements. Now I do not pretend to know what Einstein's research programme consisted of, nor do I want to say that the information can be had with ease. As opposed to Bohr, who presented his results in a way that is perfectly adapted to the methodology of Lakatos (and was criticised for this type of presentation by physicists and philosophers), Einstein in 1905 gave only very few hints about his intentions. Yet it is safe to assume that the three papers of 1905 and the later statistical papers were all part of a single and unspecified research programme (I shall call this programme *E*). This should be kept in mind when comparing Lorentz and Einstein.

---

[1] See Zahar [1973].

4 Zahar compares *L*′ and *R*. This is unhistorical. Lorentz did not hold *L*′, he held *L*. This can be seen both from his pronouncements and from his actual research. This research was designed to develop an alternative to the theory of Hertz where the field remains attached to matter. Lorentz wanted to show that electromagnetic phenomena in moving bodies can be explained on the basis of a variety of processes, all of them taking place in accordance with Maxwell's equations *in vacuo*, referred to an all-pervading aether at rest. Only part of this research programme was superseded by *R* in 1915. The remainder was eventually superseded by the quantum theory which cannot be reconciled with *R*. We must admit that the most comprehensive research programme in physics that existed around the turn of the century, *viz*. the programme of Lorentz, was never superseded in its entirety by another single research programme, but broke up into a variety of mutually irreconcilable programmes *unless* we assume that the life-work of Einstein was guided by a programme as comprehensive as that of Lorentz and unless we succeed in reconstructing its hard core. Neither Zahar, nor anyone else has done this. As a result, Zahar's paper is incomplete at a most decisive point, and its main judgements cannot be accepted as they stand.

5 For example, Zahar says that *R* superseded *L*′ (combined with Newton's theory of gravitation, *G*) in 1915. This is without interest. Nobody held *L*′. (It is of course true that various researchers restricted themselves to tiny parts of *L* and and were content with examining their properties. But it is very doubtful that we should take *them* as a measure of what does and what does not belong to a research programme.) Lorentz and his supporters held *L*, and *L* was never superseded by *R*. It *may* have been superseded by *E*, but *E* is an unknown entity.

In 1905 the situation was even more extreme. Zahar shows, to my mind completely convincingly, that *L*′, while in trouble in 1905, was not yet superseded by *R* (as available in 1905) and that it was rational to pursue it further. Replacing *L*′ by *L* considerably strengthens the case for Lorentz. For *R* (in 1905) not only does not supersede *L*, *it is a step back*, a considerable reduction in content, *without any comparable gain*. By 1905 *L* had led to the following novel facts ('novel' in the novel sense explained by Zahar which seems eminently reasonable to me): Fresnel coefficient (1892/95); experiments of Röntgen (1892), Eichenwald (1903), Blondlot (1901) and Wilson (1905) which went against Hertz and for Lorentz (*L* was essential in their explanation which could not have been achieved by *L*′ alone); dispersion formula (1892); Zeeman effect (1897); constancy of *c* and Lorentz contraction (1892/1909); Peltier effect, law of Wiedmann-Frantz and other consequences of the electron theory of metals (1904/5). Some of these facts were eventually absorbed into *R*, though only in a purely phenomenological way (phenomenological electrodynamics of Minkowski), others were taken over as 'postulates', in a simplified form, and without any further explanation, while still others had to wait for about twenty years until they were finally absorbed into the quantum theory.

6 The situation supports a pet theory of mine according to which (*i*) the transition from a research programme P to a newer and better programme Q in the same domain almost always involves a loss of content which (*ii*) is not noticed as scientists and historians use a $P' \subset P$ which has indeed much smaller content than Q. P' does not occur in history, it is a logical construct, suggested by Q (it usually can be 'derived from' Q) and by the belief in increase of content.

In the case discussed by Zahar the actual rivals are L and R. L has larger content than R, but it is not remembered. What *is* remembered is L' which follows from R and seems to have smaller content than R.

My criticism applies to Lakatos's and Zahar's fascinating analysis of the Copernican revolution[1] as well. The research programme Copernicus wants to overthrow (which I shall call A) contains a cosmology, a dynamics as well as certain methods for calculating the paths of the planets P_A. His own research programme, K, contains a new and unified method for calculating the paths of planets, P_K. Lakatos and Zahar compare P_A and P_K only. They write: 'Ptolemy's programme always dealt with facts *post hoc* and hence always degenerated. Copernicus's model of the *Commentariolus* constituted dramatic progress'. The 'dramatic progress' turns into stagnation if we compare A and K which are the research-programmes at issue (all dynamical problems have to be taken care of *post hoc*; the same is true of optical problems). Concentrating on P_A and P_K also means adopting the instrumentalism of the Ptolemaians which Copernicus tried to overcome.

7 Let us assume that L' and R are the correct rivals. Can we *now* accept the results of Zahar's analysis? I do not think we can.

R supersedes L' and G in 1915 says Zahar. This means that all the still useful facts of L' plus G are repeated by R while R provides some facts not produced by L'. One of the 'facts' of Newton's theory of gravitation taken in conjunction with L' is the long-range stability of the planetary system. To show this in R it is not sufficient to establish continuity between classical laws and the laws of R as Zahar seems to believe. One must also show that the neglected magnitudes are not related to each other in ways that may eventually lead to major discrepancies and the calculation must be pushed to magnitudes of the same order as the corresponding classical calculations, or to magnitudes of an even higher order. No such calculations were available in 1915.

8 Zahar also wants to show that those who preferred Einstein to Lorentz between 1905 and 1915 did so for 'objective' reasons. Neither he nor Lakatos has ever revealed what it is that makes a reason 'objective'. Lakatos offers standards for the comparative evaluation of research programmes, and he *calls* the standards 'objective'. But not in a single line does he explain why they are more objective than, say, the standards of Kuhn, or of Polanyi. I suspect he has the following rationale: a judgement is 'objective' if it is based on standards which lead to the same results in the same circumstances and which are themselves the outcome of 'objective' procedures. He has also specified the procedures that lead to the selection of the right standards. In all these steps objectivity *arises* as the result of a rigorous training that reinforces some reactions and suppresses others. Even

---

[1] Lakatos and Zahar [1975].

the 'objectivity' of a contradiction such as 'I am sad and I am not sad' can be guaranteed only if the customary dull interpretation of the logician is carefully separated from more imaginative readings of the phrase. We realise, then, that the word 'objective' as used by Lakatos and Zahar either refers to some choice piece of frozen thought which the reader is supposed to share with the author, or functions as an instrument for the freezing of still volatile ideas. The frozen thought to which Zahar appeals for his present purpose is not in agreement with the spirit of the methodology of research-programmes.

9 Zahar wants to show that those who accepted Einstein in 1905 could quote 'objective' reasons in their favour.

Now the most important 'objective' reasons the methodology of research programmes offers are of course progressive changes. To obtain them, one has to wait. Altogether the methodology of research programmes judges the evolution of a programme *over a period of time*, it does not judge its shape *at a particular time*. The things one does while the programme develops are done without 'objective' reasons (though there may be plenty of reasons of a different kind). Zahar's attempt to find 'objective' reasons even for this period is in conflict with the methodology of research programmes.

This becomes very clear when we look at the criterion he wants to use for excluding the aether: a research programme containing an element that has no function at time *t* is at that time inferior to a research programme that lacks the element, but gives exactly the same predictions; *temporary ad hocness* is to be frowned upon. Using the criterion he assumes that the aether had become superfluous by 1905 and that *L'* and *R* (1905) were observationally equivalent. This is not true. Even in 1905 (or in 1906, 1907, . . .) equivalence had been established for equilibrium states only: if contractions and dilatations of length, time, changes of mass are the results of an interaction between the aether and the molecules embedded in it, then any change of velocity relative to the aether will lead to oscillations. Relativity, on the other hand, knows of no such oscillations. It was only in 1936 that this consequence of *L'* was tested, and refuted by Wood, Tomlinson and Essen. We see: Zahar's criterion is neither acceptable, nor properly applied. He has not given us any objective reasons for the early acceptance of the theory of relativity.

10 To sum up: of Zahar's theses (*i*) and (*ii*) may be accepted but (*iii*) and (*iv*) must be rejected.[1] A more realistic account is needed of the rival research programmes. Accepting the methodology of Lakatos we must renounce our wish to give 'objective' reasons for *all* important scientific actions. Still, we have to admit that Zahar's account has superseded all the earlier accounts of the Lorentz-Einstein issue.

P. K. FEYERABEND
University of California, Berkeley

REFERENCES

LAKATOS, I. and ZAHAR, E. G. [1975]: 'Why did Copernicus Supersede Ptolemy?', in R. S. Westman (*ed.*): *The Copernican Acheivement.*

ZAHAR [1973]: 'Why did Einstein's Programme Supersede Lorentz's?' *BJPS*, **24**.

---

[1] Cp. *above*, p. 25, 1.

ON LORENTZ'S METHODOLOGY *

Elie Zahar, in his recent paper [1973], presents an attempt to use Lakatos's methodology of 'research programmes' to analyse the work of Einstein and Lorentz. An attempt to interpret or to explain episodes in the history of science within some larger context using an imaginative conceptual framework can result in exciting and provocative studies. However, a necessary prerequisite to this type of study is the presentation of relevant scientific literature within its proper historical context, and to take account of its internal structure. This Zahar has not done, with the result that his study becomes at best misleading, and his criticisms of the researches of others becomes empty. One reason may be that Zahar's application of the methodology of research programmes seems intended to follow the form proposed by Lakatos in his [1970]. However, Zahar takes account only of Lakatos's first rule of procedure for the use of historical case studies[1]—'(1) *one gives a rational reconstruction* [according to the methodology of research programmes]'—but not of the second one—'(2) *one tries to compare this rational reconstruction with actual history and to criticize one's rational reconstruction for lack of historicity....*'

Zahar states that his goal in his [1973] is:[2]

> ... I propose to refute *all* the charges of *ad hocness* which have been levelled at the *L.F.C.* [Lorentz-Fitzgerald contraction hypothesis] and show that Lorentz's programme progressed until after 1905.

Zahar describes Lorentz's 'ether programme' as a progressive research programme, the 'hard core' of which is 'Maxwell's equations for the electromagnetic field [and] Newton's laws of motion,' the Galilean transformation and the Lorentz force equation, and an heuristic which 'arises from the overall *metaphysical* principle that all physical phenomena are governed by actions transmitted by the ether.'[3]

Then, consonant with the methodology of research programmes, Zahar considers Lorentz's ether programme as consisting of 'three consecutive theories', T_1, T_2, T_3, a series which exhibits a progressive theoretical problem shift. According to Zahar, the theory T_1 is that proposed in Lorentz's [1892*a*]; theory T_2 is that sketched out in Lorentz's [1892*b*] and then presented in detail in Lorentz's [1895]; and theory T_3 is supposed to be the one presented in Lorentz's famous [1904]. Zahar gives the following description of T_1, T_2 and T_3:[4]

---

* This work was written while I was on sabbatical leave for 1972–73 as a National Endowment for the Humanities Fellow in the Department of Physics at Harvard University.

[1] Lakatos [1970], p. 138; italics in original.

[2] Zahar [1973], p. 99; italics in original.

[3] *Ibid.*, p. 100; italics in original. According to Zahar [1973], pp. 99–100 (and Lakatos):

> The hard core consists of assumptions which, by methodological decision, as it were, are kept unfalsified. Each theory in the programme is a conjunction of, on the one hand, the hard core and, on the other, of auxiliary hypotheses to which the modus tollens is directed whenever anomalies arise. A programme also has a heuristic which consists of a set of suggestions and hints which govern the construction or modification of the auxiliary hypotheses. The heuristic, which sets out a research policy, is less rigid than the hard core.

[4] Zahar [1973], p. 100; italics in original.

> T_1 consists of the hard core as defined above together with the (tacit!) assumptions (*i*) that moving clocks are not retarded and (*ii*) that material rods are not shortened by their motion through the ether.
>
> T_2 is obtained from T_1 by substituting the *L.F.C.* for assumption (*ii*). According to the *L.F.C.* a body moving through the ether with velocity $\vec{v}$ is shortened by the factor $\sqrt{1-v^2/c^2}$.
>
> T_3 is the conjunction of the hard core, of the *L.F.C.* and of the assumption, that, contrary to (*i*), clocks moving with velocity $\vec{v}$ are retarded by the factor $\sqrt{1-v^2/c^2}$.

He then restates his goal as:[1]

> I claim that both the shift from T_1 to T_2 and that from T_2 to T_3 were non *ad hoc*. This implies in particular that the introduction of the *L.F.C.* which took Lorentz from T_1 to T_2 was not an *ad hoc* manoeuvre.

This is essential for Zahar to prove, because if it were not the case, then Zahar would have to consider Michelson's result as 'an obstacle for Lorentz's programme';[2] thus, Zahar would have trouble arguing that the shift from T_1 to T_2 is a progressive problem shift, rather than the beginning of a new research programme. The reason is that the *L.F.C.*, stated as dimensions in the direction of motion shortened by the amount $\sqrt{1-v^2/c^2}$ (or to second order in v/c by the amount $1-v^2/(2c^2)$) is not a plausible auxiliary hypothesis because it is not in 'the spirit of the heuristic of the programme.'[3] According to Zahar, the *L.F.C.*, tagged in this way onto the hard core, would have to be considered as *ad* hoc_3.[4]

Zahar's argument for the non *ad hocness* of the *L.F.C.* turns on two points: one which concerns methodology; the other involves, as I shall demonstrate, a convenient distortion of historical facts which enables Zahar to fit Lorentz's work of 1892–1904 into the structure of the methodology of research programmes. The point of methodology is:[5]

> *Ad hocness* in research programmes is defined not as a property of an isolated hypothesis but as a relation between two consecutive theories. A theory is said to be *ad* hoc_1 if it has no novel consequences as compared with its predecessor. It is *ad* hoc_2 if none of its novel predictions have been actually 'verified'; for one reason or another the experiment in question may not have been carried out, or—much worse—an experiment devised to test a novel prediction may have yielded a negative result. Finally the theory is said to be *ad* hoc_3 if it is obtained from its predecessor through a modification of the auxiliary hypotheses which does not accord with the spirit of the heuristic of the programme.

He then defines what he means by a 'novel fact':[6]

> *A fact will be considered novel with respect to a given hypothesis if it did not belong to the problem-situation which governed the construction of the hypothesis.*

It will become evident that these redefinitions of '*ad hoc*' and 'novelty' are contrived merely to support Zahar's contention that the *L.F.C.* is not *ad hoc*, and, therefore, that the shift from T_1 to T_2 fits into the mold defined by the methodology of research programmes. In fact, according to Henri Poincaré's 'logical (*i.e.* 'rational')' judgement of what constitutes *ad hoc* statements (to be discussed below), Zahar's notion of *ad hoc* is itself *ad hoc*.

---

[1] *Loc. cit.;* italics in original.

[2] *Ibid.*, p. 107.

[3] *Ibid.*, p. 101.

[4] *Loc. cit.*

[5] *Loc. cit.*

[6] *Ibid.*, p. 103; italics in original.

Zahar's strategy is to argue that Lorentz added to T_1 an hypothesis H_2,[1] which is not *ad hoc* relative to T_1, *i.e.* is a generalisation of a statement already contained in T_1, and, moreover, H_2 does not pertain to Michelson's result. The conjunction of H_2 with the hard core of T_1 is the theory T_2. Then, according to Zahar, Lorentz *deduced* the *L.F.C.* from T_2. The *L.F.C.* is then considered to be a novel fact relative to H_2. This leads us to the second point in Zahar's argument —the one which concerns a distortion of historical fact. Zahar states it in two parts:[2]

> (Z_a): Lorentz deduced the *L.F.C.* from a deeper theory, namely from what I call the Molecular Forces Hypothesis (hereafter referred to as the *M.F.H.*) and which can be loosely formulated as follows: 'Molecular forces behave and transform like electromagnetic forces.' Moreover, in his deduction of the *L.F.C.*, Lorentz made use of his famous transformation, as is clearly indicated by the following passage from his [1895]:
>
> 'For, if we now understand by S_1 and S_2 not, as formerly, two systems of charged particles but two systems of molecules—the second at rest and the first moving with velocity v in the direction of the axis of x- between the dimensions of which the relation subsists as previously stated; and if we assume that in both systems the x-components of the forces are the same, while the y- and z-components differ from one another by the factor $\sqrt{1-v^2/c^2}$, then it is clear that the forces in S_1 are in equilibrium whenever they are so in S_2... The displacement would naturally bring about this disposition of the molecules of its own accord and thus effect a shortening in the direction of motion in the proportion of 1 to $\sqrt{1-v^2/c^2}$ ***in accordance with the formulae given in the above-mentioned paragraph.***'
>
> I further maintain that for anybody prepared to accept the assumption of an ether at rest, the *M.F.H.* is a plausible auxiliary hypothesis which introduces no alien elements into Lorentz's programme. Putting it more objectively, the theory T_2 proposed by Lorentz is non *ad hoc*$_3$, because the *M.F.H.* is structured in accordance with the heuristic of the ether programme, which requires that physical phenomena be explained in terms of actions propagated in the ether.
>
> (Z_b): Moreover, ***the M.F.H. arose out of considerations which had nothing to do with Michelson's experiment.*** The *M.F.H.* arose out of mathematical considerations pertaining to the transformation properties of Maxwell's equations. ***Hence Michelson's null result is a novel fact relative to the M.F.H.!*** The *M.F.H.* is consequently non *ad hoc*$_2$; it constituted both theoretical and empirical progress.

---

[1] H_2 will turn out to be Lorentz's hypothesis that molecular forces transform like electromagnetic forces, *i.e.* the Lorentz force.

[2] *Ibid.*, pp. 106–7. The quotation of Lorentz is from the portion of the *Versuch* [1895] that pertains to Michelson's results and translated in *P.R.C.*, p. 7. The italics are Zahar's. The passage from Lorentz's [1895] in which the Michelson and Morley experiment is discussed is translated in *P.R.C.*, pp. 3–7. All page references will be to *P.R.C.*

Zahar then attempts to buttress the line of argument presented in (Z_a) and (Z_b) by drawing upon Lorentz's [1892*a*]:[1]

> (Z_c): In Lorentz's [1892*a*] there is no mention of the 'crucial' experiment first performed by Michelson in 1881, then repeated by Michelson and Morley in 1887. This should not surprise us once we have realised that the Lorentz transformation was originally used as a mathematical device of which Lorentz gave no physical interpretation, in much the same way as we might use the expression *ict* without attaching any physical meaning to the multiplication of the distance *ct* by $\sqrt{-1}$.

To comment systematically on Zahar's analysis, it is thus necessary to analyse (Z_c) and the statements which follow, and then return to (Z_a) and (Z_b). The majority of my critical comments will be directed toward Zahar's arguments for the non *ad hocness* of the *L.F.C.*

Zahar's characterisation of the foundations of Lorentz's theory is incorrect. Newton's three laws of motion were subsequently found not to be an integral part of the theory. Lorentz, in his [1895], deduced that Newton's third law was inconsistent with the fundamental equations of the electromagnetic theory.[2] Furthermore, it is quite ahistorical to consider Maxwell's equations but not the ether as part of the hard core. The reason is that for Lorentz the Maxwell equations had no meaning without the ether. Lorentz considered the ether as a real quantity, even though its physical effects might never be detected directly. Aside from the fact that the ether constituted an integral part of electromagnetic theory *c.* 1900, it is clearly evident from even a cursory study of Lorentz's papers on electromagnetic theory that his 'research policy' was guided by the Maxwell equations and the Lorentz force equation, which were inseparable from the ether —in fact, the Maxwell equations were supposed to be the means by which one could calculate the state of the ether.

In (Z_c) is the comment that Lorentz did not mention Michelson's 'crucial' experiment in his [1892*a*]. This is not an oversight on Lorentz's part; indeed, Lorentz was very concerned about not being able to account for Michelson's result by use of his electromagnetic theory. In a letter to Lord Rayleigh on 18 August 1892, shortly after completing his [1892*a*], Lorentz wrote:[3]

> Fresnel's hypothesis [of a stagnant aether] taken conjointly with his [partial dragging] coefficient $1-1/n^2$, would serve admirably to account for all the observed phenomena were it not for the interferential experiment of Mr Michelson, which has, as you know, been repeated after I published my remarks on its original form, and which seems decidedly to contradict Fresnel's views. *I am totally at a loss to clear away this contradiction, and yet I believe if we were to abandon Fresnel's theory, we should have no adequate theory at all,* the conditions which Mr Stokes has imposed on the movement of aether being irreconcilable to each other.
>
> *Can there be some point in the theory of Mr Michelson's experiment which has as yet been overlooked?*

---

[1] Zahar [1973], pp. 111–12.

[2] Lorentz [1895], pp. 24 ff. For further discussion of the incompatibility of Newton's third law with Lorentz's electromagnetic theory see Miller [1973].

[3] Quoted from K. Schaffner [1972], p. 103. Italics added.

This information contradicts Zahar's suggestion that Lorentz in 1887 might already have thought of the contraction hypothesis, 'but would have considered such a simplistic contraction hypothesis unacceptable'.[1] Simplistic indeed, and such a hypothesis, moreover, would have been considered as *ad hoc* according to anyone's interpretation of the term. Lorentz was too good a physicist to put forth such a hypothesis without at least being able to support it with a plausibility argument; it appears that none existed in 1887.

Zahar, on the other hand, implies that Lorentz waited until he could deduce the *L.F.C.* from another auxiliary hypothesis. However, in the light of Lorentz's letter to Lord Rayleigh, this discussion is quite beside the point.

To comment upon the next sentence in (Z_0) we must discuss briefly a certain mathematical transformation used by Lorentz in [1892*a*]. Zahar is correct in noting that a version of what later became known as the Lorentz transformation was introduced by Lorentz in [1892*a*] as a mathematical aid in solving the inhomogeneous second order partial differential equations for the electromagnetic field quantities in an inertial reference system.[2] The transformation is:[3]

$$x' = \gamma x_r \tag{1}$$

$$y' = y_r \tag{2}$$

$$z' = z_r \tag{3}$$

$$t' = t - \gamma^2 \frac{v}{c^2} x_r \tag{4}$$

where $x_r = x - vt$, $y_r = y$, $z_r = z$, $t_r = t$; $\gamma = 1/\sqrt{1-v^2/c^2}$; the coordinates x, y, z, t are measured relative to a reference system S fixed in the ether; x', y', z', t' are measured relative to a reference system Q' (Zahar calls it S'), which has no physical meaning; and x_r, y_r, z_r, t_r refer to a real inertial system S_r. In fact, in Q' the speed of light is not c to all orders in v/c. This can easily be shown as follows. The transformations inverse to (1)–(4) that relate S and Q' are:

$$x = \gamma x' + vt' \tag{5}$$

$$y = y' \tag{6}$$

$$z = z' \tag{7}$$

$$t = t' + \gamma \frac{v}{c^2} x' \tag{8}$$

To facilitate the remainder of this proof, I resort to the following device: in S we have

$$x^2 + y^2 + z^2 = c^2 t^2 \tag{9}$$

which means that the speed of light is c in S. Transforming (9) to Q' by use of (5)–(8) yields

$$x'^2 + y'^2 + z'^2 = c^2(1 - v^2/c^2)t'^2 \tag{10}$$

thus, in Q' the speed of light is

$$c\sqrt{1 - v^2/c^2}, \tag{11}$$

---

[1] Zahar [1973], p. 115.

[2] *Ibid.*, p. 112.

[3] See Lorentz [1892*a*], p. 297. Zahar quotes these equations on his p. 112.

and not c. Lorentz was aware of this shortcoming of the system Q', although, of course, not by the method I outlined above. Lorentz, as a result of inspecting the functional form of the fields obtained by use of transforming the field equations from S_r to Q', could not have failed to see that the electromagnetic field quantities propagate with the velocity $\sqrt{c^2-v^2}$, and, therefore, may have no physical interpretation. However, to first order in v/c these fields do propagate with the velocity c. Lorentz gave two reasons for making this approximation: it facilitates subsequent calculations; and, more important, it permitted him to prove '*un théorème général*'[1]—namely, that to first order in v/c the electromagnetic field quantities for an harmonically bound microscopic charged particle (as he referred to the electron in his [1892*a*]) have the same form in S_r as they do in S. To this order of accuracy the spatial coordinates in Q' become identical to those in S_r, and the time coordinate becomes what Lorentz was to define in his [1895] as the 'local time'.[2] The '*théorème général*' can be considered as an earlier version of the 'theorem of corresponding states'[3] from Lorentz's [1895].

In summary: Q' is a fictitious coordinate system which can be considered as at rest in the ether and in which the speed of light is c only to order v/c. Q' is obtained from S_r by extension of the coordinates along the x_r-axis, *i.e.* by a change of units.

In the second sentence of (Z_c) Zahar asserts that Lorentz accounted for Michelson's result by using Q'. Zahar then amplifies this statement:[4]

> (Z_d): However, soon after writing his [1892*a*], Lorentz realised that the transformation equations lent themselves to an interpretation which he immediately set out in his [1892*b*] and which he then expounded in greater detail in his [1895].

An analysis of Lorentz's [1892*b*] and [1895] will throw light upon the erroneous statements in (Z_a), (Z_b), (Z_c), (Z_d) and elsewhere in Zahar's paper. First I shall outline *Lorentz's* methodology in [1892*b*] and [1895], and then compare this to Zahar's Lorentz—the one who supposedly did his physics according to the methodology of research programmes.

Lorentz began [1892*b*] by discussing his analysis given in detail in his [1886] of Stokes' ether theory and Michelson's experiment. He then discussed the Michelson and Morley experiment of 1887. *Then*, by means of a purely geometrical argument, *i.e.* one based upon the Newtonian addition law for velocities, Lorentz showed that their null result could be accounted for to second order in v/c by positing that the length of the interferometer arm in the direction of motion be shortened by an amount $1-v^2/(2c^2)$. This subsequently became known as the *L.F.C.* Thus far, Lorentz has made no mention of what Zahar refers to as the 'Molecular Forces Hypothesis' (see (Z_a) and (Z_b), hereafter referred to as *M.F.H.*). Clearly, the *L.F.C.*, as first proposed by Lorentz in [1892*b*], was a physics of desperation—he says as much:[5]

> (L_a): This experiment [Michelson and Morley] has been puzzling me for a long time, and in the end I have been able to think of only one means of reconciling its result with Fresnel's theory.

---

[1] Lorentz [1892*a*], p. 304. [2] Lorentz [1895], p. 50. [3] *Ibid.*, p. 85.

[4] Zahar [1973], p. 112.

[5] Lorentz [1892*b*], p. 221. See also the letter of Lorentz to Lord Rayleigh quoted above.

If one wishes to place this calculation within a hypothetico-deductive scheme, then one is forced to say that Lorentz deduced the *L.F.C.* from a law of mechanics.

Lorentz continues in [1892*b*] to give what is clearly a plausibility argument for the *L.F.C.*:[1]

> (L_b): Now, some such change in the length of the arms in Michelson's first experiment and in the dimensions of the slab in the second one is so far as I can see, not inconceivable. What determines the size and shape of a solid body? Evidently the intensity of the molecular forces; any cause which would alter the latter would also influence the shape and dimensions. Nowadays we may safely assume that electric and magnetic forces act by means of the intervention of the ether. It is not far-fetched to suppose the same to be true of the molecular forces. But then it may make all the difference whether the line joining two material particles shifting together through the ether, lies parallel or crosswise to the direction of that shift. It is easily seen that an influence of the order of p/V is not to be expected, but an influence of the order of p^2/V^2 is not excluded and that is precisely what we need.

Lorentz emphasises that the hypothesis of the *L.F.C.* is 'not inconceivable'. He then posits the hypothesis H', which is that 'it is not far-fetched to suppose' that the molecular forces which after all determine the shape of a body 'act by intervention of the ether'. Thus, at this point, Lorentz's only assertion about molecular forces is that they do not propagate instantaneously.

Lorentz concludes this passage by emphasising the necessity of proving that molecular forces are affected by motion through the ether in such a way that they influence the shape of an object in motion by at least a term of second order in (v/c). He then asserts that H' cannot be tested:[2]

> (L_c): Since the nature of the molecular forces is entirely unknown to us, it is impossible to test the hypothesis.

What is needed is the very bold hypothesis which we referred to as H_2—namely, that the 'influences of the motion of ponderable matter on electric and magnetic forces' is the same for molecular forces.[3] Thus, Lorentz's next step is to calculate the Lorentz force in S_r. Lorentz did this soon after the publication of [1892*a*]. His method was to make use of the fictitious system Q' so as to transform a problem of electrodynamics into one of electrostatics, the latter class of problems are easier to solve. As evidenced in [1892*b*], he obtained a result accurate only to second order in (v/c):

$$F'_x = F_x \tag{12}$$

$$F'_y = \left(1+\frac{v^2}{2c^2}\right)F_y \tag{13}$$

$$F'_z = \left(1+\frac{v^2}{2c^2}\right)F_z \tag{14}$$

where primed quantities refer to the fictitious system Q'. According to H_2, molecular forces transform like the Lorentz force, and, therefore, relative to S

---

[1] Lorentz [1892*b*], p. 221.

[2] *Loc. cit.*

[3] *Ibid.*, pp. 221–2.

propagate through the ether with the speed of light in vacuum. This is Zahar's *M.F.H.* Equations (*12*)–(*14*) indicate that a system in equilibrium under molecular forces in S_r should also be in equilibrium in Q', which is taken as at rest in the ether. Lorentz then argues that this can be so only if dimensions in S_r in the direction of motion are shortened by the amount $(1-v^2/(2c^2))$. Lorentz obtained this result by writing (*1*)–(*3*) in the approximate form

$$x_r = \left(1-\frac{v^2}{2c^2}\right)x' \qquad (15)$$

$$y_r = y' \qquad (16)$$

$$z_r = z' \qquad (17)$$

in order to compare dimensions in S_r and Q'. Thus, putting forth the *M.F.H.* allowed Lorentz to give theoretical support for the *L.F.C.* As if to emphasize this point, Lorentz goes on to assert:[1]

> (L_d): One may not of course attach much importance to this result; the application to molecular forces of what was found to hold for electric forces is too venturesome for that.

Lorentz continues by asserting that the *L.F.C. might not be the correct explanation* of the Michelson and Morley experiment:[2]

> (L_e): Besides, even if one would do so, the question would still remain whether the earth's motion shortens the dimensions in one direction, as assumed above, or lengthens those in directions perpendicular to the first, which would answer the purpose equally well.

In his [1895] Lorentz applies the electromagnetic theory that he had developed at length in his [1892*a*] to problems involving optical and electrical phenomena in moving bodies. Lorentz discussed two different transformations that he had developed in this treatise for two different sets of problems. The set of transformations:[3]

$$x' = \gamma x_r \qquad (18)$$

$$y' = y_r \qquad (19)$$

$$z' = z_r \qquad (20)$$

$$t' = t_r \qquad (21)$$

permitted Lorentz to prove that to order v/c electrostatic processes are unaffected by the earth's motion and also could be used to transform certain problems in electrodynamics to those in electrostatics. The primed coordinates refer to a fictitious system S', which bears some similarity to the Q' system from the [1892*a*]. S' is the electron's instantaneous rest system, in which the electron is dilated along the x'-axis by an amount $1/\sqrt{1-v^2/c^2}$. Using (*18*)–(*21*), Lorentz was able to relate the Lorentz forces in S' and S_r:

---

[1] *Ibid.*, p. 223.

[2] *Loc. cit.*

[3] Lorentz [1895], pp. 35–9. For a discussion of these equations see Miller [1973], esp. parts 3 and 4.

$$F_x = F'_x \tag{22}$$

$$F_y = \sqrt{1-v^2/c^2}F'_y \tag{23}$$

$$F_z = \sqrt{1-v^2/c^2}F'_z \tag{24}$$

which is a more exact version of (*12*)–(*14*).

The other system of transformation is:

$$x_r = x-vt \tag{25}$$

$$y_r = y \tag{26}$$

$$z_r = z \tag{27}$$

$$t_L = t-\frac{v}{c^2}x \tag{28}$$

where t_L is the local time and the spatial portion (*25*)–(*27*) pertains to the Galilean transformations. This system relates S_r and S. Lorentz postulated the mathematical concept of the local time coordinate to prove the covariance to first order in v/c of the Maxwell equations in charge free space or within a substance that is neutral, not magnetic and not a conductor.[1] Thus, equations (*25*)–(*28*) are the set of transformations that Lorentz used in the theorem of corresponding states. After having accounted systematically for the result of all first order optical experiments, Lorentz, in the final chapter of the [1895], discussed three experiments accurate to second order in v/c which could not be accounted for 'without further ado'. One of them is the Michelson and Morley experiment. Lorentz's methodology in dealing with the Michelson and Morley experiment parallels that given in his [1892*b*], but contains more details. An outline is sufficient for my purposes: After using the Newtonian addition law for velocities to discuss the Michelson and Morley experiment, Lorentz put forth the *L.F.C.* to explain their null result. However, as in [1892*b*], he goes on to assert that this account is not unique. For, if the dimensions of an object parallel to its direction of motion are changed in the ratio 1 to $1+\delta$, and those normal to its motion in the ratio 1 to $1+\epsilon$, then it follows that the Michelson and Morley result can be accounted for to second order in v/c if ϵ and δ satisfy the equation

$$\epsilon-\delta = \frac{v^2}{2c^2}. \tag{29}$$

Equation (*29*) is one equation for two unknowns; hence, either ϵ or δ would remain undetermined. The *L.F.C.* corresponds to $\epsilon = 0$. Then Lorentz tries to make 'this hypothesis' (one which he confesses to be 'surprising') more acceptable, and employs a characteristically modest tone that shows how tentative and dependent on a string of assumptions his suggestion really is.[2]

> (L_f): Surprising as this hypothesis may appear at first sight, yet we shall have to admit that it is by no means far-fetched, as soon as we assume that molecular forces are also transmitted through the ether, like the electric and

---

[1] Lorentz [1895], '*Abschnitt* V'. See also Miller [1973], esp. part 4.

[2] *P.R.C.*, pp. 5–6.

> magnetic forces of which we are able at the present time to make this assertion definitely. If they are so transmitted, the translation will very probably affect the action between two molecules or atoms in a manner resembling the attraction or repulsion between charged particles.

This is the *M.F.H.*, which Lorentz proposed *after* putting forth the *L.F.C.* That Lorentz put forth the *M.F.H.* to assist him in making plausible the *L.F.C.* is clear from his next statement in the [1895]:[1]

> (L_g): From the theoretical side, therefore, there would be no objection to the hypothesis.

Then, at the *end* of this section, Lorentz presents yet another plausibility argument—full of honestly stated and interconnected assumptions—in support of contraction only of the dimensions along the direction of motion. This final section, quoted in its entirety, is:[2]

> (L_h): It is worth noticing that we are led to just the same changes of dimensions as have been presumed above if we, *firstly*, without taking molecular movement into consideration, assume that in a solid body left to itself the forces, attractions or repulsions, acting upon any molecule maintain one another in equilibrium, and, *secondly*—though to be sure, there is no reason for doing so—if we apply to these molecular forces the law which in another place we deduced for electrostatic actions. For if we now understand by S_1 and S_2 not, as formerly, two systems of charged particles, but two systems of molecules—the second at rest and the first moving with a velocity v in the direction of the axis of x—between the dimensions of which the relationship subsists as previously stated; and if we assume that in both systems the x components of the forces are the same, while the y and z components differ from one another by the factor $\sqrt{1-v^2/c^2}$, then it is clear that the forces in S_1 will be in equilibrium whenever they are so in S_2. If therefore S_2 is the state of equilibrium of a solid body at rest, then the molecules in S_1 have precisely those positions in which they can persist under the influence of translation. The displacement would naturally bring about this disposition of the molecules of its own accord, and thus effect a shortening in the direction of motion in the proportion of 1 to $\sqrt{1-v^2/c^2}$, in accordance with the formulæ given in the above-mentioned paragraph. This leads to the values
>
> $$\delta = -\tfrac{1}{2}\frac{v^2}{c^2}, \quad \epsilon = 0$$
>
> in agreement with (1).
>
> In reality the molecules of a body are not at rest, but in every 'state of equilibrium' there is a stationary movement. What influence this circumstance may have in the phenomenon which we have been considering is a question which we do not here touch upon; in any case the experiments of Michelson and Morley, in consequence of unavoidable errors of observation, afford considerable latitude for the values of δ and ϵ.

This argument is similar to the one given in his [1892*b*]. S_1 and S_2 are what I refer to as S_r and S', respectively. Using (*18*)–(*20*) and (*22*)–(*24*), Lorentz argues that it is not unreasonable to propose that $\epsilon = 0$.[1]

---

[1] *Ibid.*, p. 6.

[2] *Ibid.* pp. 6–7.

In the final paragraph Lorentz once again emphasises that $\epsilon = 0$ may not be the correct choice; however, it is the only one which he can argue for on the basis of his electromagnetic theory, even though the argument makes use of a fictitious coordinate system. Lorentz still considers the *L.F.C.* as a physics of desperation, put forth to preserve his electromagnetic theory.

With this development, we can now easily show that Zahar's arguments for the non *ad hocness* of the *L.F.C.* are specious.

In (Z_a) Zahar asserts that 'Lorentz deduced the *L.F.C.* from a deeper theory', namely from what he calls the *M.F.H.* However, as we have just seen, Lorentz did no such thing. Moreover, the *M.F.H.* does *not*, contrary to Zahar, constitute a 'theory about molecular forces'.[2] For example, Lorentz's suggestion that molecular forces transform like electromagnetic forces does not even give us information about the functional form of molecular forces, as a theory about molecular forces would require.[3] Nor does Lorentz furnish such a theory elsewhere. See also (L_e), wherein Lorentz asserts that the 'nature of molecular forces is entirely unknown'. Thus, an objective rendering of Lorentz's [1892*b*] and [1895] makes manifest the incorrectness of Zahar's statement (Z_b), thereby also severely weakening his (Z_a), to which I now turn.

In (Z_a) Zahar, in order to support his claims about the *M.F.H.*, quotes at length from Lorentz's [1895]; however, this passage is from the *final* portion of Lorentz's lengthy work [1895]—compare (Z_a) with (L_h). As I emphasised, Lorentz in (L_h) first gave a plausibility argument for $\epsilon = 0$, and then concluded on a note of uncertainty as to the veracity of the *L.F.C.* Thus, Zahar is quite incorrect when he states:[4]

> *Lorentz Derived the L.F.C. from the M.F.H. . . . the Michelson–Morley Experiment Lends Dramatic Support to the M.F.H.*

Moreover, the 'famous transformation' that Zahar mentions in (Z_a) is not what can be considered, even with hindsight, as the 'Lorentz transformation' of 1895. As I mentioned previously, there were two sets of transformations in the [1895]: equations (*18*)–(*21*); and equations (*25*)–(*28*). The former set was used in the passage from the [1895] quoted in (Z_a). The latter set is what can, in retrospect, be considered by us as (*c.* 1895) the 'Lorentz transformations'. They are the set Lorentz used to prove the theorem of corresponding states. Zahar is obviously unaware of this fact. Indeed, Zahar, in his discussions of Lorentz's work of 1892 and 1895, in particular of the 1895 theorem of corresponding states,[5] gives the impression that Lorentz used the set of transformations from S_r to Q' from his [1892*a*] in the *Versuch* [1895]—this is incorrect. Thus, also incorrect is the import of Zahar's statement that:[6]

> *the First Version of Lorentz's Theory of Corresponding States Arises out of the Realistic Interpretation of the Lorentz Transformation whose Origins were Purely Mathematical.*

---

[1] For further discussion see Miller [1973].

[2] Zahar [1973], p. 111.

[3] Another example: To assert merely that the gravitational force in Newton's theory of gravitation transforms like a vector gives no information about its functional form.

[4] Zahar [1973], p. 114; italics in original.

[5] Zahar [1973], pp. 111 ff.

[6] *Ibid.*, p. 111; italics in original.

In summary: It is clear from Lorentz's papers [1892*b*] and [1895] that his methodology was to posit the *L.F.C.* specifically to account for the Michelson and Morley experiment. Moreover, he obtained the *L.F.C.* from an argument based upon the Newtonian addition law for velocities. Then, Lorentz posited the *M.F.H.* for the purpose of constructing a plausibility argument in support of the *L.F.C.* that involved the use of certain concepts from his electromagnetic theory—namely, the transformation equations from S_r to Q' in [1892*b*], and from S_r to S' in [1895].[1] Lorentz stated explicitly in [1892*b*] and [1895] that the *L.F.C.* might not be the proper manner of accounting for the result of Michelson and Morley. In addition, Lorentz argued (in [1892*b*], p. 223, and [1895], p. 6) that the *L.F.C.* can very probably not be directly observed. Thus, to save his electromagnetic theory, Lorentz put himself into the position of arguing on the basis of a set of transformation equations to a fictitious coordinate system and a string of interconnected assumptions for the plausibility of an effect that probably could not be directly observed. Thus, Lorentz saw fit to put forth this 'surprising' hypothesis in a very modest manner.

These are the hard historical facts upon which Zahar must try to superimpose the methodology of research programmes. It will not do to distort history at one's convenience and then to criticise the analyses of other historians and philosophers of science with the aid of suitably redefined terms such as 'novelty' and '*ad hoc*'.

Indeed, Zahar conveniently forgets to mention that, until 1904, the majority of physicists did consider the *L.F.C.* as *ad hoc*; in particular, Poincaré, whose criticism was of great concern to Lorentz,[2] and Lorentz himself. Moreover, Poincaré's reasons for considering the *L.F.C.* as *ad hoc* as it appeared in the *Versuch* [1895] (but less so as it was presented in Lorentz's [1904]) were quite logical: In the *Versuch* the *L.F.C.* was presented to account for one experiment, and, furthermore, it did not fit into the body of the treatise, which discussed a theory to account systematically only for first order effects. On the other hand, in Lorentz's [1904], the *L.F.C.* in conjunction with other hypotheses put forth in this paper could account for more experiments than that of Michelson and Morley, *e.g.* those of Trouton and Noble, Rayleigh, Brace and Kaufmann. Moreover, in Lorentz's [1904] the *L.F.C.* fits into the body of the theory presented, which claimed to be able to account for effects to all orders in v/c. Zahar may have omitted this well-known episode in the history of science because it was not consonant with fitting Lorentz's work into the methodology of research programmes.

Although Poincaré's judgement of the *L.F.C.* as *ad hoc* is one based upon logic—or as Lakatos would say, in terms of Poincaré's logical decision—judgements of *ad hocness* do not always follow from purely logical arguments. Holton argues for this point in his [1969]. He asserts that there are also instances where the judgement of a hypothesis as *ad hoc* is not made by narrowly conceived ideas of 'logic' alone. The weight of Holton's supporting arguments for these claims seem to be of great concern to Zahar as a proponent of the methodology of research programmes. This is evidenced in Zahar's paper, wherein Zahar permits Grünbaum to easily dispose of Popper's case for the *ad hocness* of the *L.F.C.*[1] and spends most of his space in an effort attempting to refute the examples cited in Holton [1969]. Whether one agrees with Holton's conjectures or not, a close study of his papers (especially his [1969]) reveals that Holton's methodology is first to present as objectively as is possible the historical facts relevant to the case under study, taking particular care *not* to distort the internal structure of the relevant literature, and only then to analyse or speculate—the reverse of Zahar's methodology.

---

[1] In fact, Lorentz, in his review paper [1903]—to which Zahar does not refer—clearly separates the *L.F.C.* from the *M.F.H.* Lorentz, in Section 62 ('The Interference Experiment of Michelson', pp. 273–4) gives an argument based upon the Newtonian addition rule for velocities, asserting that the Michelson and Morley experiment could be accounted for if the dimensions of a body in motion are changed according to my equation (*29*). *Then*, in Section 64 ('Present Status of the Theory') Lorentz asserts that this hypothesis 'becomes somewhat more intelligible if we put forth that [molecular forces behave like the Lorentz force, *i.e.* the *M.F.H.*]' (p. 277).

[2] See Lorentz [1904], pp. 12–13. See also Miller [1973], esp. part 5.2.

As to Zahar's criticisms of Holton's [1969] themselves, they can be more quickly disposed of. The sentences from Holton [1969] to which Zahar directs what he claims to be factual criticism are:[2]

> (H_a): This saving *Hülfshypothese* [the *L.F.C.*] is introduced completely *ad hoc* . . . *No explicit comment is made which connects this assumed shrinkage with the Lorentz transformations in their still primitive form, as published earlier in the book* . . .
>
> (H_b): The contraction hypothesis when it was made was clearly and quite blatantly *ad hoc*—or, if one prefers to use the *patois* of the laboratory, *ingeniously cooked up for the narrow purpose which it was to serve* . . .
>
> (H_c): The important point to note is that '*ad hoc*' is not an absolute but a *relativistic term*. Postulate 1 and 2 [Einstein's two postulates in his [1905]] may be said to have been introduced *ad hoc* with respect to the Relativity Theory of 1905 as a whole . . . But these postulates were *not ad hoc with respect to the Michelson experiment*, for they were not specifically imagined in order to account for its results . . .

Previously, I discussed how Lorentz, in the [1895], put forth a plausibility argument for the *L.F.C.*, using the transformations from the real inertial system S_r to the fictitious system S'. I then pointed out Zahar's misconception of this point. I also noted that the transformation from S_r to S' cannot be considered as the 1895 version of what were subsequently called the Lorentz transformations. Thus, Holton's (H_a) is correct.

Holton's (H_b) is a restatement of previously reported opinion *c.* 1900 concerning the *L.F.C.*, *e.g.* that of Poincaré. It is a historical fact which cannot be made to vanish simply by redefining the notion of *ad hoc*.

Zahar's claim to refutation of (H_c) is in the following footnote:[3]

> I cannot make head or tail of Holton's sentence 'Postulates 1 and 2 may be said to have been introduced *ad hoc* with respect to the Relativity Theory

---

[1] Zahar [1973], p. 104. More precisely, according to Zahar Grünbaum showed that the *L.F.C.* is not *ad hoc*$_1$, as was claimed by Popper. However, if one wishes to analyze the status of the *L.F.C. c.* 1900, a goal to which Zahar aspires, then Grünbaum's argument is quite beside the point. The reason is that Grünbaum's argument turns upon the Kennedy–Thorndike experiment of 1932.

[2] Zahar [1973], pp. 105–6; italics by Zahar.

[3] Footnote 1 in *ibid.*, p. 106.

> of 1905 as a whole'. How can Einstein's two postulates, which constitute Special Relativity Theory or at any rate are part of Relativity Theory, be *ad hoc* with respect to Relativity Theory?

Ad hoc with respect to x means not derivable from x. The (H_c) sentence means that Postulates I and II are not derived from relativity theory, but are postulated prior to (and so become a part of) relativity theory. Evidence for this assertion is as follows: (*i*) Autobiographical remarks by Einstein; (*ii*) raising a 'Vermutung' to a postulate; (*iii*) argument that goes as follows—a careful reading of the scientific literature relevant to electron physics in 1905–10 reveals that Einstein's theory was widely considered in the community of physicists to be a generalisation of Lorentz's theory of the electron presented in Lorentz's [1904]. Indeed, after 1905 the nomenclature Lorentz–Einstein theory came into being. The prediction of this theory that was considered of prime importance was for the so-called transverse mass

$$m_T = \frac{m}{\sqrt{1-v^2/c^2}}.$$

(As is well known, Einstein, due to an unfortunate choice for the definition of force in his [1905], arrived at a different expression for m_T. This oversight was corrected by Planck in 1906. Thus, Einstein's and Lorentz's predictions for m_T are symbolically identical.) Holton's point in (H_c) is as follows: From the result that $m_T = m/\sqrt{1-v^2/c^2}$ one is not necessarily led to Einstein's two postulates of relativity, since m_T could also be deduced from Lorentz's theory of the electron of 1904. This leads me to another of Zahar's incorrect statements:[1]

> (Z_e): The philosophical significance of the Theory of Corresponding States is that it could, as Poincaré showed, easily be turned into a theory observationally equivalent to Special Relativity.

Zahar also makes the claim[2] (repeats an old mistake) that Einstein's and Lorentz's theories are observationally equivalent. Special relativity is not observationally equivalent to Lorentz's theory of 1904. The reason is that a theory of light containing Lorentz's ether cannot account *exactly* for the optical Doppler effect, nor for observations of stellar aberration.[3] Neither are the two theories observationally equivalent concerning the *L.F.C.* The reason is that Lorentz's contraction is basically different from that of Einstein. In fact, Lorentz, by applying the classical notion of congruence to the behaviour of rods as implied by his theory, convinced himself that the *L.F.C.* could not be observed.[4] Lorentz did not realise the importance of the comparison of lengths using light signals. Einstein and Poincaré simultaneously and independently discovered this, although their reasoning on this point differed. Even if Lorentz had realized this point before the others did, the two contractions would still not be observationally equivalent. The reason is that in Lorentz's theory the class of ether-fixed reference systems is not equivalent to the class of inertial systems, whereas in Einstein's theory of relativity of 1905 this asymmetry does not exist. This lack of symmetry in Lorentz's theory between different sets of space-time coordinates results in a lack of symmetry in the observation of the contraction. Furthermore, the inclusion by Lorentz of true lengths and times, quantities which remain unknown to the observer in an inertial system, renders unclear the meaning of the contraction of a rod at rest relative to an inertial system S_r^1 by an observer in an inertial system S_r^2.

---

[1] Zahar [1973], p. 116.

[2] *Ibid.*, p. 123.

[3] This point will be discussed fully in a forthcoming work by the author. For some brief remarks the reader should consult Møller [1952], esp. pp. 8–10; Drude [1959], esp. the first footnote on p. 475.

[4] Lorentz [1892*b*], p. 223 and [1895], p. 6 of *P.R.C.*

I should now like to comment upon Zahar's concluding section of his Part I, entitled '*The Rationality of Lorentz's Pursuing his own Programme after 1905*'.[1]

Contrary to what Zahar asserts,[2] there was no 'Relativity Programme' in the scientific community from 1905 to about 1910, because Einstein's work had not yet been disentangled from Lorentz's—there existed for most scientists only one theory, the Lorentz–Einstein theory. Moreover, the Lorentz–Einstein theory was part of a research effort that can be called the perfection of the electromagnetic world-picture, formally announced in 1900 by Wilhelm Wien.[3] The intent of this research effort was to formulate an electromagnetic field-theoretical description of the electron and, subsequently, of all matter. It was a reductionistic programme that attempted to reduce the laws of mechanics to those of electromagnetism, *i.e.* to Lorentz's fundamental equations.[4] Thus, contrary to Zahar's characterisation of Lorentz's theory, by 1904 Newton's third law of motion was no longer considered to be an integral part of this theory. I believe that Zahar, by referring to Lorentz's theory as the '*Theory of Corresponding States*',[5] has missed this point. An immediate result of Zahar's oversight is that Lorentz's theory of the electron is made to appear to be more universal than it actually was.[6]

In summary: It was quite 'rational' for Lorentz to continue to adhere to his theory of the electron until 1910. However, Lorentz maintained this viewpoint to the end of his life (1928). Can this fact be explained entirely on 'rational' grounds? Or, does it really serve the philosophy of science so narrowly to construe the idea of 'rational' or 'logical' scientific thinking by actual creative research scientists that any deviation from it which occurred historically must be falsified or explained away?

---

[1] Zahar [1973], pp. 122–23; italics in original.

[2] *Ibid.*, p. 122.

[3] Wien [1900]. See Miller [1973], esp. part 2, for further details.

[4] See Miller [1973] for further discussion. One of the reasons for the emergence of the electromagnetic world-picture was a presumed incompatibility between the laws of mechanics and those of electromagnetism: The laws of electromagnetism, *i.e.* those of Lorentz's theory of [1892*a*] and [1895], transformed according to the modified Galilean set of transformations (25)–(28). On the other hand, the laws of mechanics transformed according to the usual Galilean transformations, *i.e.* (25)–(27) with $t_r = t$ replacing (28). The former theory with its set of transformation equations could account systematically for the results of optical and electrical experiments to order v/c. The latter theory could not. In view of the successes of Lorentz's theory *c.* 1900, *vis-à-vis* the lack of success of mechanical models for electromagnetism, many physicists believed that the best way to resolve the tension between the laws of electromagnetism and those of mechanics was to reduce mechanics to electromagnetism. Thus, in light of these facts, we must consider Zahar's statement that 'Michelson's experiment did not *refute* the conjunction of Newton's laws and Maxwell's equations' ([1973], p. 115; italics in original) to be true only up to 1900.

[5] Zahar [1973], p. 116; italics in original.

[6] Incidentally, there were at least two other major theories of the electron to compete with the Lorentz–Einstein theory: Langevin-Bucherer and that of Abraham. See Miller [1973] for further discussion.

I would finally comment on certain of Zahar's assertions concerning the work of Poincaré:[1]

> But despite this, one hesitates to count Poincaré among classical physicists such as Maxwell and Lorentz. Poincaré was the first scientist to recognise the group character of the transformation equations and probably also the first clearly to enunciate a physical principle of relativity. (Einstein is supposed to have carefully read 'Science and Hypothesis' before 1905.) Whatever the case may be, Poincaré showed, using Lorentz's own approach, how the Theory of Corresponding States could be made observationally equivalent to Special Relativity.

Zahar's statement that 'Poincaré was the first scientist to recognise the group character of the' Lorentz transformations can be contested. The reason is that this statement is also put forth by Einstein in his [1905].[2] Thus, we have another case in the history of science of simultaneous but independent discovery.

Zahar's statement that Poincaré was the 'first clearly to enunciate a physical principle of relativity. (Einstein is supposed to have carefully read *Science and Hypothesis* before 1905)' implies that Poincaré discussed in his book *Science and Hypothesis*, published in 1902, the principle of relativity as posited again in 1904. This is not true. In *Science and Hypothesis* Poincaré discussed the 'principle of relative motion'[3] in *classical mechanics*. This principle is a restatement of the Newtonian principle of relativity (*i.e.* Newton's Corollary V). In the final chapter, 'Electrodynamics', Poincaré mentioned the term 'principle of relativity' (in common use *c.* 1902), but in reference to the laws of geometry.[4] Furthermore, it has been repeatedly shown in recent publications that Poincaré's principle of relativity was basically different from Einstein's. Poincaré's contained Lorentz's ether; Einstein's did not. Poincaré's was to be one of the principles of a reductionistic programme; Einstein's embraced all branches of physics, and represented one of the principles of a viewpoint that did not attempt to reduce one branch of physics to another.

Thus I fear that Zahar is on the wrong track—the reason is that there is no such thing as what he refers to as Lorentz's T_3: for Lorentz always considered the local time as merely a mathematical coordinate transformation.

ARTHUR I. MILLER
Lowell Technological Institute, Lowell, Massachusetts

REFERENCES

DRUDE, P. [1959]: *The Theory of Optics*, translated from the German edition of 1900.

EINSTEIN, A. [1905]: 'Zur Elektrodynamik bewegter Körper,' *Ann. d. Phys.* **17**, 891–921, in *The Principle of Relativity*, translated by W. Perrett and G. B. Jeffery. This reprint volume is referred to as *P.R.C.* All references are to *P.R.C.*

HOLTON, G. [1969]: 'Einstein, Michelson and the "Crucial" Experiment,' *Isis*, **60**, pp. 133–97.

---

[1] Zahar [1973], p. 120.

[2] Einstein [1905], p. 51.

[3] Poincaré [1902], p. 111. See Miller [1973], esp. part 5, for a discussion of Poincaré [1902].

[4] *Ibid.*, p. 244.

LAKATOS, I. [1970]: 'Falsification and the Methodology of Scientific Research Programmes', in I. Lakatos and A. E. Musgrave (*eds.*): *Criticism and the Growth of Knowledge*, pp. 91–195.

LORENTZ, H. A. [1886]: 'De l'influence du mouvement de la terre sur les phénomènes lumineux', *Versl. Kon. Akad. Wetensch. Amsterdam*, **2**, p. 297 (1886); in *Collected Papers*, **4**, p. 153.
Unless indicated otherwise, all references to Lorentz's work will be to the *Collected Papers*.

LORENTZ, H. A. [1892*a*]: 'La théorie électromagnétique de Maxwell et son application aux corps mouvants,' *Arch. néerl.* **25**, p. 363 (1892); in *Collected Papers*, **2**, p. 164.

LORENTZ, H. A. [1892*b*]: 'The relative motion of the earth and the ether', *Versl. Kon. Akad. Wetensch. Amsterdam*, **1**, p. 74 (1892), in *Collected Papers*, **4**, p. 219.

LORENTZ, H. A. [1895]: *Versuch einer Theorie der elektrischen und optischen Erscheinungen in betwegten Körpern* (1st *ed.* 1895; 2nd *ed.* 1906). All references will be to 1906.

LORENTZ, H. A. [1903]: 'Weiterbildung der Maxwellschen Theorie. Elektronentheorie', *Encykl. math. Wiss.*, **2**, Art. 14, pp. 145–280 (1904; received December 1903).

LORENTZ, H. A. [1904]: 'Electromagnetic phenomena in a system moving with any velocity smaller than that of light', *Proc. Roy. Acad. Amsterdam*, **6**, p. 809 (1904), in *P.R.C.*, pp. 11–34.

MILLER, A. I. [1973]: 'A Study of Henri Poincaré's "Sur la Dynamique de l'Electron" ', *Archive for History of Exact Sciences*, **10**, pp. 207–328.

MØLLER, C. [1952]: *The Theory of Relativity*.

POINCARÉ, H. [1902]: *Science and Hypothesis*.

SCHAFFNER, K. [1972]: *Nineteenth-Century Ether Theories*.

WIEN, W. [1900]: 'Über die Möglichkeit einer elektromagnetischen Begründung der Mechanik', *Recueil de travaux offerts par les auteurs à H. A. Lorentz*, pp. 96–107.

ZAHAR, E. [1973]: 'Why did Einstein's Programme Supersede Lorentz's? (I),' *British Journal for the Philosophy of Science*, **24**, pp. 95–123.

EINSTEIN VERSUS LORENTZ: RESEARCH PROGRAMMES AND THE LOGIC OF COMPARATIVE THEORY EVALUATION [*]

Introduction.

1 *Historical Background.*
  1.1 *The Lorentz Electron Theory.*
  1.2 *The Einstein Theory of Special Relativity.*
2 *The Methodology of Comparing Competing Theories.*
  2.1 *The Concept of an* Ad Hoc *Hypothesis.*
  2.2 *The Formalism of a Logic of Theory Competition.*
3 *'Why Did Einstein's Programme Supersede Lorentz's?'.*
  3.1 *The Simplicity of Einstein's Theory.*
  3.2 *Intertheoretic Consideration Favouring Einstein over Lorentz.*

Introduction.

Elie Zahar's recent two-part paper[1] raises a most interesting set of historical and methodological questions. Utilising, with some minor divergences, the approach developed by Imre Lakatos[1] out of the clash between the philosophy of scientific growth of Sir Karl Popper[2] and Thomas Kuhn's analysis of scientific revolutions,[3] Zahar has examined in a stimulating fashion the comparative development of H. A. Lorentz's and Albert Einstein's theories of the electrodynamics of moving bodies. In so doing he has called attention in a concrete manner to that unit of scientific growth—the Lakatos research programme—which is more general than a scientific theory and less 'occult' than Kuhn's nevertheless most fruitful concept of a 'paradigm'. Zahar has argued, correctly I think, and following in the Popperian tradition of Agassi, Watkins[4] and Feyerabend[5], for the important role of metaphysics in the incipient stages of theory genesis and growth. He has proposed an account of the guiding heuristics of the Lorentzian and Einsteinian research programmes which, somewhat controversially, unifies the Einsteinian approach involved in the special theory with that employed in the genesis of the general theory of relativity. He has also provided us with suggestive analyses of the concepts of *ad hocness* and novelty.

---

[*] Grateful acknowledgement is made to the U.S. National Science Foundation for earlier support of research on late-aether theories and the special theory of relativity.

[1] Zahar [1973].

Zahar has, however, committed several serious historical errors in his account. His philosophical analyses also raise some paradoxes which can in part be resolved by a more accurate history, but which also seem to me to require a more developed or fine-structure analysis of the logic of comparative theory evaluation for a full resolution. I shall contend below that Zahar has incorrectly argued for the progressive non-*ad hoc* character of Lorentz's theory after 1904. I shall also suggest that Zahar has misunderstood the second postulate of Einstein's special theory of relativity, and that this has both falsely led him to assume the postulate introduced certain imperfections or incoherencies into the theory, and to perform unnecessary gyrations concerning Einstein's thinking in the genesis of the special theory. I also believe Zahar has ignored central scientific and metascientific features of Einstein's contribution in the special theory, namely his reanalysis of the time-concept and simultaneity, and the considerable degree of simplicity which such a reanalysis allowed in the electrodynamics of moving bodies. I shall also argue that insufficient attention to intertheoretic consideration is responsible for the paradox that, though Einstein's theory was accepted by many prominent scientists by 1910–12, Zahar can only fully rationalise its acceptance after 1915.

I shall also disagree with some of the major factors which Zahar conceives of as relevant to the genesis of the special theory and shall provide an alternative account which is coherent with my above mentioned point of view, and also with the available historical documentation.

1 *Historical Background.*

1.1 *The Lorentz Electron Theory.*

(*a*) *Lorentz's Theorem of Corresponding States.*

There are several points which should be noted in connection with both theoretical and experimental aspects of H. A. Lorentz's theory of electrons. One interesting verbal slip is to be found in Zahar's characterisation of Lorentz's '*theorem* of corresponding states' as a '*theory* of corresponding states'. The slip s significant since it is associated with what I take to be a misconstrual of the thrust of Lorentz's research programme in the years 1892–1910. Though Lorentz does not use the term 'theorem' in his [1904], he does employ it in his works of [1895], [1899], and [1909]. The change in locution is important inasmuch as Lorentz's approach begins with his axioms or his fundamental equations postulating the existence of charged electrons ('ions' before 1900), a stationary aether, and the micro-behaviour of the electromagnetic field connecting the electrons. (Zahar is somewhat misleading when he suggests that the 'hard core' of Lorentz's theory consists of 'Maxwell's equation for the electromagnetic field . . .',[1] since one important contribution of Lorentz's theory even in its earliest forms ([1892] and [1895]) was to derive Maxwell's macroscopic equations from his own microscopic equations which possessed a strong analogy with Maxwell's equations.) The fundamental equations of Lorentz's theory:

---

[1] Lakatos [1968], [1970].

[2] Popper [1959], [1962].

[3] Kuhn [1962].

[4] Agassi [1964], Watkins [1958].

[5] Feyerabend [1964].

$$\begin{aligned} &\mathit{div}\,\mathbf{D} = \rho \\ &\mathit{div}\,\mathbf{H} = 0 \\ &\mathit{curl}\,\mathbf{H} = \frac{1}{c}\left(\frac{\partial \mathbf{D}}{\partial t} + \rho \mathbf{v}\right) \\ &\mathit{curl}\,\mathbf{D} = -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t} \end{aligned}$$

$$\mathbf{F} = \mathbf{D} + \frac{1}{c}(\mathbf{v} \times \mathbf{H})$$

served as a basis for a theory of electromagnetic processes in moving bodies (*i.e.* bodies in motion through the aether). As I noted in my [1969],[2] Lorentz proceeded in a way that Einstein later would have called 'constructive'.[3] He began with the above fundamental equations of the electron theory, introduced the classical or 'Galilean' velocity addition formula or transformation for electrons moving with the system in motion through the aether, and *after* applying the classical transformation $x' = x - vt$, the transformation equations bearing Lorentz's name were added. The Lorentz research programme for the years 1892–1910 becomes fully intelligible only if one understands that the goal of Lorentz's endeavours was to provide 'Lorentz transformation equations' which would allow him to *prove* his *theorem* of corresponding states. The transformation equations evolved from 1895 to 1904 in several steps, changing the form of particular transformations in minor ways (*e.g.*, the sequence of transformations for 'local time'[4]) and adding new transformations (*e.g.*, the Lorentz ([1904]) transformation for the electric moment **P**), in an attempt to obtain the theorem of corresponding states which would cover recently suggested or performed experiments. Lorentz ([1899]) cited Liénard's variant of the Michelson interferometer experiment in which the rays pass through a solid or liquid dielectric as motivation for exploratory research on transformations of the second order of v/c. In his

---

[1] Zahar [1973], p. 100.

[2] Schaffner [1969], p. 507.

[3] See Einstein [1949], p. 53 and Einstein [1934], pp. 74–5, for a discussion of the term 'constructive'. I discuss the role of 'constructive' theories versus 'principle-type' theories in my [1974*a*].

[4] The evolution of the concepts of time and space in Lorentz, Poincaré, and Einstein (as regards the special theory of relativity) is covered in detail in my [1974*a*]. Those portions of that paper which touch on the genesis of Einstein's theory are also presented below in section 1·2(*c*).

[1904] Lorentz referred to Rayleigh's ([1902]) and Brace's ([1904]) attempts to determine if a Lorentz–Fitzgerald contraction might cause stresses resulting in a transparent body becoming doubly refracting, and also mentioned Trouton and Noble's ([1903]) experiment to ascertain the torque acting on a charged capacitor as a result of the capacitor's motion through the aether. Without further modifications, the Lorentz electron theory would have required positive results for these experiments; with further modifications in the transformation equations, the theorem of corresponding states could be saved.

It is important to also note explicitly (Zahar does not) that Lorentz did *not* believe that the theorem of corresponding states *should* be tantamount to a 'principle of relativity' in either Poincaré's ([1900], [1902], [1904], [1905] and [1906]) or Einstein's ([1905] and [1907]) senses. In his [1904] Lorentz stated that he wished 'to show by means of certain fundamental assumptions and without neglecting terms of one order of magnitude or another, that *many* electromagnetic actions are *entirely independent* of the motion of the systems' (my italics). The 'many' is significant for it was because of his commitment to the causal efficacy of the aether that Lorentz could not accept either Poincaré's or Einstein's approaches, prior to about 1910. In his [1909] Lorentz wrote:

> I cannot but regard the ether, which can be the seat of an electromagnetic field with its energy and its vibrations as endowed with a certain degree of substantiality, however different it may be from all ordinary matter. In this line of thought it seems natural not to assume at starting that it can never make any difference whether a body moves through the ether or not, and to measure distances and lengths of time by means of rods and clocks having a fixed position relatively to the ether.[1]

Lorentz ([1909]) added further that he had not yet accepted the important modifications proposed independently by Poincaré ([1906]) and Einstein ([1905*b*]) for charge density and velocity addition transformations. Lorentz noted:

> Whereas I have not been able to obtain for the equations referred to moving axes *exactly* the same form as for those which apply to a stationary system, Einstein has accomplished this by means of a system of new variables slightly different from those which I have introduced. I have not availed myself of his substitutions, only because the formula are rather complicated and look somewhat artificial, unless one deduces them from the principle of relativity itself.[2]

The mathematical formalism is accordingly conditioned by Lorentz's aether metaphysics.

Zahar suggests that Lorentz could not achieve the correct transformations because he would have had to have 'developed a general kinematical framework'.[3] This does not seem to be correct, both because of Lorentz's own remarks cited immediately above, and because Poincaré, proceeding *within* the Lorentzian approach, *was* able to arrive at the correct transformations, but only because he (Poincaré) possessed a strong commitment to a 'principle of relativity'. (Poincaré's relativity principle is conceptually different from Einstein's inasmuch it is still aether-theoretic. Nonetheless as a methodological tool it was sufficient to generate the formally correct transformation equations for charge density and velocity addition.)

---

[1] Lorentz [1909], p. 230.

[2] [1909], p. 230.

[3] Zahar [1973], p. 119.

(*b*) *The Lorentz–Fitzgerald Contraction and the Molecular Forces Hypothesis.*

In his attempt to show that the Lorentz *research programme* was not *ad hoc* in any one of his senses, Zahar's [1973] places much emphasis on the role of what he terms the Lorentz 'molecular forces hypothesis', or *M.F.H.* for short. The role of the *M.F.H.* in alleviating the *ad hoc* character of the Lorentz–Fitzgerald contraction hypothesis (*L.F.C.*) has been noted several times in the literature,[1] but Zahar provides an interestingly novel perspective on the *M.F.H.* by suggesting that *this hypothesis* is itself *ad hoc* in *none* of his senses. Though I shall postpone a detailed methodological analysis of the notion of *ad hocness* until Section 2 below, it will be useful to comment briefly on Zahar's analysis of *ad hoc* at this point, and then to consider the origin, meaning, and *evolution* of the Lorentz *M.F.H.* in the light of this account.

Zahar ([1973]) proposes that there are three senses of *ad hoc*. He states that '*ad hocness* in research programmes is defined not as a property of an isolated hypothesis but as a relation between two consecutive theories'[2], but he is either not entirely faithful to this stipulation, or else hypotheses can be considered *ad hoc* in some derivitive sense, since he does discuss the non *ad hoc* character of the *M.F.H.*, for example.[3] Though I shall postpone my analysis of *ad hocness* (and of Zahar's concept of 'novelty') until section 2, it will be useful to introduce Zahar's three senses of *ad hoc* in the context of the *M.F.H.* Zahar has suggested the following analyses of *ad hocness*:[4]

ad hoc$_1$: A theory is said to be *ad hoc*$_1$ if it has no novel consequences as compared with its predecessor.

ad hoc$_2$: [A theory] . . . is *ad hoc*$_2$ if none of its novel predictions have been actually 'verified'; for one reason or another the experiment in question may not have been carried out, or—much worse—an experiment devised to test a novel prediction may have yielded a negative result.

ad hoc$_3$: [A] . . . theory is said to be *ad hoc*$_3$ if it is obtained from its predecessor through a modification of its auxiliary hypotheses which does not accord with the spirit of the heuristic of the programme.

The notion of 'novelty' which appears in all three definiens above is characterised by Zahar in the following terms:

> A fact will be considered novel with respect to a given hypothesis [sic] if it did not belong to the problem-situation which governed the construction of the hypothesis.[5]

Let us now consider the *M.F.H.*. Zahar formulates the *M.F.H.* in admittedly somewhat loose phraseology as the hypothesis that 'Molecular forces behave and transform like electromagnetic forces'.[6] Lorentz in his first enunciation of the hypothesis was somewhat more precise proposing that the molecular forces under consideration were 'mutual attractions and repulsions' between systems of material points, these systems constituting what we interpret as solid bodies. Lorentz ([1892*b*]) also noted that 'Nowadays we may safely assume that electric and magnetic forces act by means of the intervention of the ether. It is not far-fetched to suppose the same to be true of the molecular forces'.

---

[1] See Møller [1952], p. 29, and Schaffner [1969], pp. 499–500.

[2] Zahar [1973], p. 101.

[3] Zahar [1973], p. 107.

[4] Zahar [1973], p. 101.

[5] Zahar [1973], p. 103.

[6] Zahar [1973], p. 106.

Zahar argues that Lorentz's theory with the *L.F.C.* (which he calls T_2) is not *ad hoc* in any one of the three senses quoted above. It is not *ad hoc*$_1$ insofar as the Kennedy–Thorndike experiment can falsify it;[1] it is not *ad hoc*$_3$ because 'Lorentz deduced the *L.F.C.* from a deeper theory involving the *M.F.H.* This latter hypothesis, Zahar contends, is not *ad hoc* because the *M.F.H.* is structured in accordance with the heuristic of the ether programme, which requires that physical phenomena be explained in terms of actions propagated in the ether'.[2] Finally the *M.F.H.* (and *a fortiori* T_2) is not *ad hoc*$_2$ because '*the M.F.H. arose out of considerations which had nothing to do with Michelson's experiment.* The *M.F.H.* arose out of mathematical considerations pertaining to the transformation properties of Maxwell's equations. *Hence Michelson's null result is a novel fact relative to the M.F.H.*!'

Let us consider the historical situation in which Lorentz first proposed the *M.F.H.* The *L.F.C.* and the *M.F.H.* were proposed in the same paper (Lorentz [1892*b*]) so it seems likely that both hypotheses were part of the same 'problem-situation'. Lorentz did *not* conjecture that the *molecular forces* would alter *in the same specific way* that electric forces were changed by motion through the aether *except* in considering the Michelson experiment. Furthermore, at the time that Lorentz suggested the *M.F.H.*, it possessed no additional known *empirically* testable consequences than the Michelson experiment. There was, as I point out below, some *theoretical* support for the *M.F.H.* Later Lorentz *further generalised* the *M.F.H.* to cover *any* forces, intermolecular, *intra*-atomic and *intra*-electronic, in systems in motion through the aether. These generalisations were, however, forced on Lorentz because of certain experiments performed in 1902–4 which would have *falsified* his theory without further modification. These experiments, those of Rayleigh ([1902]), Brace ([1904]) and Trouton and Noble ([1903]), were described briefly above.

It is important to examine these tactics of Lorentz in his ([1904]) generalisation of the *M.F.H.*, since I believe that it will show what Lakatos ([1970]) has termed degenerating problemshift. I also think it will *a fortiori* demonstrate that the Lorentz research programme was engaging in retrogressive *ad hoc* moves from 1904 through 1910.

The *M.F.H.* possessed some plausibility in its 1892 and 1895 forms inasmuch as the molecular forces could be thought of as electromagnetic in nature, and the force transformation expressions developed by Lorentz would apply, derivatively, to such molecular forces, and would entail the *L.F.C.* This form of the *M.F.H.* depended then on molecular forces being *reducible* to electromagnetic forces for its non *ad hoc* character. This *reduction thesis* which was a rather general theoretical thesis probably possessed sufficient independent support in the late nineteenth-century so that it would alleviate, if not completely eliminate, the *ad hoc* character of the *M.F.H.*

---

[1] Zahar thus accepts Grünbaum's arguments set forth in his [1959] and [1963] concerning the falsification of the *L.F.C.* (without a time dilation hypothesis) by the Kennedy–Thorndike experiment.

[2] Zahar [1973], p. 106.

In his [1904] Lorentz proposed what he felt at the time to be two additional and independent hypotheses to be conjoined to the earlier ([1892*a*], [1895] and [1899]) forms of his theory of electrons. One hypothesis was that of the contractile electron (*C.E.*). Whereas before 1904 Lorentz had applied the *L.F.C.* only macroscopically, because of the Rayleigh, Brace and Trouton-Noble experiments, Lorentz realised that he had to *extend the contraction to individual electrons.* Without a further hypothesis concerning the sub-structure of the electron, and Lorentz did *not* provide one in his [1904], the *M.F.H.* hypothesis will not entail the *C.E.* The *C.E.* hypothesis is, however, independently testable by determinations of the deflections of 'radium-rays' in electric and magnetic fields, and was so tested with vacillating results prior to about 1908.[1]

The second hypothesis which Lorentz added in his [1904] was the generalisation of the *M.F.H.* Lorentz wrote:[2]

> I shall suppose *that the forces between uncharged particles, as well as those between such particles and electrons, are influenced by a translation in quite the same way as the electric forces in an electrostatic system.*

This hypothesis was not *explicitly* applied to the *C.E. per se* in his [1904] though it is implicitly applied. In Lorentz's ([1899] and [1904]) transformations, a factor ϵ (in [1899]) or l (in [1904]) appears. In his [1904] Lorentz determines the value of l as equal to unity by utilising the generalised *M.F.H.* together with Newton's second law of motion and (indirectly) his postulated transformation for the dielectric displacement of the aether.[3] He then asserts that:

> We are therefore led to suppose *that the influence of a translation on the dimensions (of the separate electrons and of a ponderable body as a whole) is confined to those that have the direction of the motion, these becoming β times smaller than they are in the state of rest.*[4]

That the electrons do *not* contract in the dimensions orthogonal to their motion and that the electrons contract only a factor of β, and not βl in the direction of motion is, then, partially based on the generalised *M.F.H.*

In his [1909] Lorentz *explicitly* applied the generalised *M.F.H.* to the substructure of electrons. Lorentz wrote:

> Either a spherical electron must be regarded as a material system between whose parts there are certain forces ensuring the constancy of its size and form, or we must simply assume this constancy as a matter of fact which we have not to analyse any further. In the first case the form, size and orientation of the moving ellipsoid will also be maintained by the action of the system of forces, provided all of them have the property expressed in our

---

[1] An important experiment by Bucherer ([1908]) confirmed the Lorentz–Einstein deformable electron. For other later experiments supporting Lorentz and Einstein see Lorentz [1915], p. 339, footnote 86.

[2] Lorentz [1904], p. 21.

[3] 'Indirectly' since the Lorentz derivation is *via* his calculation of the electromagnetic momentum in a moving system, and this is obtained from his transformation for the dielectric displacement **D**′.

[4] Lorentz [1904], p. 28.

> relation . . . [the generalised *M.F.H.*]. In the other case we may rest content with simply admitting for the moving ellipsoid, without any further discussion, the ellipsoidal form with the smaller axis in the line of translation.[1]

There were however serious reservations about the satisfactory applicability of such a generalised *M.F.H.* to electrons. First, it was shown by Poincaré ([1906]) that the contractile electron could be considered a stable entity only if a definitively *non*-electromagnetic counter-pressure were invoked. To extend the *M.F.H.* to cover *this* type of force would violate the reduction thesis (of the *M.F.H.* to electromagnetic forces) which provided the plausibility (and independent support) for the original form of the *M.F.H.* If, on the other hand, one takes Lorentz's second 'case' quoted above, the generalised form of the *M.F.H.* becomes a 'sleeping partner' in the now spurious 'derivation' of the specific form of the contractile electron, and possesses no independent testability. It would accordingly be *ad hoc*, proposed only to account for the *prima facie* falsifying experiments of 1902–4.

To sum up this rather complex evolution then, it would appear that the Lorentz theory of 1904, which is usually taken as the definitive form of his theory and that which is to be compared with Einstein's special theory of relativity, is flawed by the incorporation of the generalised form of the *M.F.H.* which is *ad hoc* in either the third sense of Zahar, inasmuch as the extension of transformation properties to definitively non-electromagnetic forces would seem to violate the 'spirit' of the Lorentz research programme, or in his second sense, insofar as a generalised *M.F.H.* not applied to electrons *per se* implied no additional novel facts not part of the problem-situation of Lorentz's 1904 theory. It is not unimportant that Einstein, in his 1917 popular exposition of relativity theory, commented on two classes of experimental support for relativity theory, and noted the following about Lorentz's contractile electron:

> . . . Lorentz was the first to introduce the hypothesis that the form of the electron experiences a contraction in the direction of motion in consequence of that motion, the contracted length being proportional to the expression $\sqrt{1-(v^2/c^2)}$. This hypothesis, which is *not justifiable by any electrodynamical facts*, supplies us then with that particular law of motion which has been confirmed with great precision in recent years (my italics).[2]

Minkowski also thought of the *L.F.C.* as *ad hoc* or 'artificial'. He wrote in his [1908*a*] that:

> H. A. Lorentz has found out the 'Relativity-theorem' and has created the Relativity postulate as a hypothesis that electrons and matter suffer contractions in consequence of their motion according to a certain law.
>
> A. Einstein has brought out the point very clearly, that this postulate is not an artificial hypothesis but is rather a new way of comprehending the time-concept. . . .[3]

In his [1908*b*] Minkowski characterised the *L.F.C.* as 'extremely fantastical' since:

---

[1] Lorentz [1909], p. 215.
[2] Einstein [1917], p. 51.
[3] Minkowski [1908*a*], p. 54.

> the contraction is not to be looked upon as a consequence of resistances in the ether, or anything of that kind, but simply as a gift from above,—as an accompanying circumstance of the circumstance of motion.[1]

In his [1905*b*], Einstein applied his derived 'Lorentz' transformations to a 'rigid sphere of radius R' to show that it would suffer a *L.F.C.*, and his derivation of the longitudinal and transverse masses of moving electrons demonstrates that he conceived the *L.F.C.* as applying at the micro-level. For Einstein such a contraction at any level was a kinematical consequence, and he did not require any hypothesis about *forces* affecting the *form* of the electron.

1.2 *The Einstein Theory of Special Relativity.*

(*a*) *Interpretation of Einstein's Light Postulate.*

Zahar has suggested that the second postulate of Einstein's theory, the light postulate, introduces a type of internal incoherence into the special theory, and that the postulate is 'thrown out with no justification whatever'.[2] In fact, arguing conditionally against Holton's [1969], Zahar suggests that *if* one is disposed to view Lorentz's programme with the *L.F.C.* as *ad hoc*, then the light postulate of the special theory of relativity is at least as *ad hoc*. Zahar writes:

> Einstein could hardly have regarded the *L.F.C.* as more *ad hoc* than his own light postulate. For, first, on Einstein's own criteria, the *S.R.T.* as presented in 1905, was far from being 'internally perfect'. It consisted of two heterogeneous parts which were on totally different levels: on the one hand a high-level, universal covariance principle and on the other a so-called light postulate which was low-level and extremely counter-intuitive.[3]

Zahar further adds:

> Unlike his first postulate (the Relativity Principle) *P1*, the Light Principle *P2* is thrown out with no justification whatever. Moreover on the face of it, *P2* runs counter to Einstein's prescription (*1*) [a principle of simplicity and internal perfection or coherence]: there seems to be no connection at all between the fundamental properties of space-time and those of light.[4]

Zahar believes this problem is resolvable only by withdrawing to Maxwell's equations, and conceiving of Maxwell's theory, and not the light postulate, as both the historical and epistemological basis of Einstein's 'second starting point—the first being the Relativity Postulate . . .'.[5]

It is my opinion that this construal of the relation of the light postulate to both the genesis and the epistemological foundations of special relativity is incorrect. I believe that Zahar's concerns rest on a conflation of two interpretations of the light postulate, one weak and classically supported by the Maxwell and Lorentz theories, and one strong, and rationalisable only in the context of Einstein's radical reanalysis of the time-concept. Zahar is not alone in conceiving of the light postulate in only its strong form, and one can find similar readings of the postulate in the writings of philosophers (Reichenbach[1]), historians (Keswani[2]), and physicists (Mermin[3]). (To the best of my knowledge, it is only Grünbaum[4] who has stressed the need to *presuppose* a reanalysis of the time-concept before a proper interpretation of the light postulate, in its strong form, can be provided.) Zahar deliberately ignores the Einstein reanalysis of synchronisation in his account, and as I shall argue below, I believe that this is an important reason for his incorrect perspective on both the genesis and the conceptual analysis of the postulates of the special theory of relativity. Let me first analyse the two interpretations of the light postulate before turning to Einstein's reconstruction of the time-concept.

---

[1] Minkowski [1908*b*], p. 81.

[2] Zahar [1973], p. 232.

[3] Zahar [1973], p. 108.

[4] Zahar [1973], p. 232.

[5] Zahar [1973], p. 233.

Zahar (and the other writers cited) interpret the light postulate as asserting the *constancy* of the velocity of light in all Galilean reference frames: 'Einstein *bluntly asserted* that the speed of light is an invariant, an assumption from which Michelson's result trivially follows.'[5] I will term this interpretation of the light postulate the constancy of light velocity hypothesis (*C.L.V.*). This hypothesis is to be sharply distinguished from a hypothesis which asserts that a *source* of light contributes no velocity from its motion to the velocity of the emitted light. The alternative to this hypothesis, which I shall term the hypothesis of the *independence* of the velocity of light from its source (*I.L.V.*), is a *ballistic* hypothesis such as suggested by Ritz and others in the early years of the twentieth-century. If a source moves with velocity v with respect to an observer, the ballistic hypothesis assigns a velocity $c+v$ to the velocity of the light emitted in the same direction of motion as the source. The *I.L.V.* hypothesis, on the other hand, assigns a velocity c to the emitted light, but this velocity c, is c *only* in the aether rest frame. For an observer in the aether rest frame, if a source moves through the aether with velocity v, the observer determines the velocity of the emitted light to be c. If the *observer* is moving through the aether, with velocity v, he determines on the basis of the *I.L.V.* hypothesis that the velocity of light emitted from any source, *whatever* the source's velocity, is $c+v$ for light moving away from the observer's direction of motion.

In his fundamental paper on the special theory of relativity, Einstein ([1905*b*]) introduced his light postulate in the following way:

> We will ... also introduce another postulate, which is only *apparently irreconcilable* with the former [principle of relativity], namely, that light is always propagated in empty space with a definite velocity *c which is independent of the state of motion of the emitting body* (my italics).[6]

After his analysis of simultaneity, a point to which I shall return at length several times below, Einstein restated the light postulate in more formal terms writing:

> Any ray of light moves in the 'stationary' system of co-ordinates with the determined velocity *c*, whether the ray be emitted by a stationary or by a moving body. Hence velocity = light path/time interval *where time interval is to be taken in the sense of the definition in* **1.** [the section developing the new analysis of time and simultaneity][7] (my italics).

---

[1] Reichenbach [1924], pp. 94–5.

[2] Keswani [1965], pp. 284–6.

[3] Mermin [1968], pp. 9–11.

[4] Grünbaum [1964].

[5] Zahar [1973], p. 108.

[6] Einstein [1905*b*], p. 38.

[7] Einstein [1905*b*], p. 41.

My interpretation of the Einstein use of the light postulate is as follows. Einstein accepts the *I.L.V.* hypothesis as the weakest physical hypothesis which he needs in conjunction with the principle of relativity to generate the (later termed) Lorentz transformation equations for space and time. The *I.L.V.* hypothesis is irreconcilable with the principle of relativity only under classical assumptions regarding simultaneity. The apparent contradiction or *paradox* is resolved by the reanalysis of synchronisation or simultaneity, which is a necessary *prior step* to providing the stronger *C.L.V.* interpretation of the light postulate.

My interpretation is supported by several lines of evidence involving both logical and historical considerations.

First, the *C.L.V.* hypothesis and the principle of relativity are not even *prima facie* contradictory. What Einstein noted as a *paradox* is intelligible only on the interpretation of the light postulate as the *I.L.V.* hypothesis. Einstein stressed this relation of apparent 'irreconcilability' and 'incompatability' in a number of his writings after 1905. For example in [1917] he wrote:

> As a result of an analysis of the physical conceptions of time and space, it became evident that *in reality there is not the least incompatability between the principle of relativity and the law of propagation of light* (Einstein's italics).[1]

Second, Einstein, in discussing the strong evidential support for the light postulate, conceived it in terms of the electromagnetic aether theories of Maxwell and Lorentz. It is important to note that Zahar, as noted above, did *not* conceive of the light postulate as having any justification. Einstein, however, wrote in 1917 concerning the light postulate that:

> The epoch-making theoretical investigations of H. A. Lorentz on the electrodynamical and optical phenomena connected with moving bodies show that experience in this domain leads conclusively to a theory of electromagnetic phenomena, of which the law of the constancy of velocity of light *in vacuo* is a necessary consequence. Prominent theoretical physicists [in the late nineteenth century] were therefore more inclined to reject the principle of relativity, in spite of the fact that no empirical data had been found which were contradictory to this principle.[2]

Zahar cites a similar passage from Einstein's later [1950] in which Einstein appeals to the 'Maxwell–Lorentz theory of electrodynamics' as the source of scientists' confidence in the light postulate,[3] but as noted above he incorrectly argues that this must mean that 'Both historically and epistemologically speaking, Einstein's second starting point—the first one being the Relativity Postulate—is not the Light Principle but the proposition: *(P3) Maxwell's equations express a law of nature*'. Zahar suggests this, in part I believe, because under his interpretation of the light postulate, the postulate has no support.

That the light postulate is entailed by Maxwell's and Lorentz's theories is further supported by Ritz's strategy in his [1908]. Ritz wished to retain the principle of relativity and also Newtonian kinematics by rejecting the light postulate. He realised, however, that this also meant rejecting Maxwell's equations and providing a modified electrodynamics.

There is a third set of considerations which I believe will both support my

---

[1] Einstein [1917], pp. 19–20.
[2] Einstein [1917], p. 19.
[3] Zahar [1973], p. 232.

I.L.V. interpretation of the light postulate. This set of considerations is based on the early writers on relativity such as Tolman ([1910]) and Pauli ([1921]).

In his excellent monograph on relativity, W. Pauli ([1921]) wrote concerning the light postulate:

> Lorentz and Poincaré had taken Maxwell's equations as the basis of their considerations. On the other hand it is absolutely essential to insist that such a fundamental theorem as the covariance law should be derivable from the simplest possible basic assumptions. The credit for having succeeded in doing just this goes to Einstein. He showed that only the following single axiom in electrodynamics need be assumed: The *velocity of light is independent of the motion of the light source*. . . . [This independence] proves to be the true essence of the old aether point of view.[1]

In 1910 R. C. Tolman felt it was important to distinguish between the 'principle' that the velocity of light is independent of its source and the stronger 'second postulate' of relativity, which asserted that 'the velocity of light in free space appears the same to all observers regardless of the relative motion of the source of light and the observer'.[2] In Tolman's view, the stronger assertion (which I have termed the *C.L.V.* hypothesis) is 'obtained by combining the first postulate of relativity with the principle that the velocity of light is independent of the velocity of the source' (the *I.L.V.* hypothesis). Tolman noted that this latter principle:

> has hitherto lacked experimental justification. It was obtained, however, as a direct consequence of the ether theory of light, which makes the velocity depend only upon the properties [*i.e.*, elasticity or electrical nature (this is an implicit reference to Maxwell's and Lorentz's theories)] of a stationary transmitting medium, and therefore, as with sound or other wave motions, independent of the velocity of the source.

I shall comment on the effect which quantum mechanics and light quanta had on both the genesis and the acceptance, by the scientific community in the early nineteen-tens, of the special theory below. It is important to note here, however, that the *I.L.V.* hypothesis is a *universal* consequence of the Maxwell and Lorentz theories *even in those areas in which the theories begin to break down because of quantum mechanical considerations.* There is good historical evidence that this fact was paramount in Einstein's mind at the time of the genesis of the special theory of relativity. Einstein had already written his [1905*a*] paper on light quanta and had defined certain limits of applicability of the Maxwell and Lorentz theories prior to developing his special theory of relativity. I shall cite Einstein further below on his belief that the special theory of relativity transcended any quantum mechanical difficulties.

In sum, my disagreement with Zahar on the status of the light postulate can be put by noting that my view is that the light postulate in its *I.L.V.* interpretation possessed considerable theoretical support in 1905, and that this support was maintained even after the realisation of quantum effects and the limited validity of Maxwell's and Lorentz's theory. Contrary to Zahar, I conceive of Einstein's theory as possessing simplicity and coherence which exceeded that of Lorentz's theory.

---

[1] Pauli [1921], p. 5.

[2] Quotations are from Tolman [1910], pp. 27, 28 and 39.

The *I.L.V.* hypothesis, however, is more important from the perspective of an analysis of the conceptual foundations of the special theory of relativity (and in the 'context of justification') than it is in the genesis of the theory. I shall argue below that Einstein initially began with the *C.L.V.* hypothesis *without* his reanalysis of the time-concept, but that this situation was deeply paradoxical. Before I turn to an account of the genesis of relativity it would be useful to indicate exactly what Einstein's reanalysis of the time-concept involved, and also to show what an impact this reanalysis possessed historically.

(*b*) *Einstein's Analysis of the Time-Concept.*

I find it significant that though Zahar does note that Einstein's theory of special relativity involved an approach which '*sorts out his kinematics* before *imposing the condition of Lorentz-covariance on all physical laws*, . . . [whereas] *Lorentz painfully struggles to arrive at a new kinematics* via electromagnetism',[1] Zahar's account of both the genesis and the reason for the acceptance of Einstein's theory says essentially nothing about the reanalysis of the time concept and the redefinition of simultaneity. This is in marked contrast to how Einstein's contemporaries felt, and it is in part due to this omission of Zahar's that I believe his perspective on both the discovery of special relativity and the supersedure of Lorentz's theory of the electrodynamics of moving bodies is fatally flawed. In this section I shall quote from several scientists who were active in electrodynamics in the early twentieth-century so as to bring out the force of Einstein's contribution. In the following section I shall develop an account of the genesis of relativity theory which is more faithful to the historical documentation than I felt Zahar's to be.

As I mentioned above, the reconciliation between the two postulates of the special theory of relativity presupposed a new understanding of time. In his 'Autobiographical Notes' Einstein wrote:

> Today everyone knows, of course, that all attempts to clarify this paradox [of the conjunction of the light postulate with the principle of relativity] satisfactorily were condemned to failure as long as the axiom of the absolute character of time, *viz.*, of simultaneity, unrecognisedly was anchored in the unconscious. Clearly to recognise this axiom and its arbitrary character really implies already the solution of the problem.[2]

Einstein's approach is too well known to require anything more than the briefest of summaries. Einstein argued that though local simultaneity determinations seemed clear, the concept of distant simultaneity was not. A light signal synchronisation procedure could be introduced so that, taking two distant points *A* and *B*, we 'establish *by definition* that the time required by light to travel from *A* to *B* equals the time it requires to travel from *B* to *A*'. (Einstein's italics in his [1905*b*]). A definitional component in the reconstruction of the concept was required because the velocity of light in a *one way* direction cannot be ascertained in any *empirical* manner prior to a determination of distant simultaneity. Since this is an Einsteinian and not a Newtonain world, such a definition is *admissible*. As Grünbaum has clearly shown, if clocks were *not* affected by their motion or if there *were* signals with a greater velocity than *c*, such a definition would not be permitted, for true and false synchronisations could be distinguished by citing these other procedures. In addition, I myself believe that the admissibility of Einstein's definition depends on the *I.L.V.* hypothesis holding. For if a Ritz ballistic hypothesis were true in nature, absolute measurements of simultaneity would be definable, as Ritz clearly realised.

---

[1] Zahar [1973], pp. 119–20.

[2] Einstein [1949], p. 53.

The definition of simultaneity which Einstein proposed in his [1905*b*] is also the most *descriptively* simple, as has been shown by Reichenbach [1927], Grünbaum [1963] and Winnie [1970]. The definition of simultaneity allows for resolution of the paradox mentioned above, but at the price of *relativising* both simultaneity determinations and length measurements, a consequence to which Einstein was swift to draw attention in his [1905*b*]. The two postulates, their contradiction thus resolved, 'suffice for the attainment of a simple and consistent theory of the electrodynamics of moving bodies based on Maxwell's theory for stationary bodies'.[1]

Einstein's approach through a reanalysis of the time-concept startled his contemporaries. Some documentation of this thesis is in order.

Max Born began his study of physics in 1901 and was interested in optics, electrodynamics, and the theory of electrons. He studied the papers of Hertz, Fitzgerald, Larmor, Lorentz, and Poincaré, among others, in a seminar which he took with Minkowski in 1905. In 1907 he had his attention directed to Einstein's papers and read them. Born has written that 'although I was quite familiar with the relativistic idea and the Lorentz transformations, Einstein's reasoning was a revelation to me'. Born further added:

> For me—and many others—the exciting feature of this paper [Einstein [1905*b*]] was not so much its simplicity and completeness, but the audacity of challenging Isaac Newton's established philosophy, the traditional concepts of space and time. That distinguishes Einstein's work from his predecessors and gives us the right to speak of Einstein's theory of relativity, in spite of Whittaker's ([1953]) different opinion.[2]

I quoted Minkowski's position on Einstein's accomplishment above, where it was noted that Einstein's fundamental advance was his new time-concept. Max Planck seems to have held a similar position, for in a famous letter of recommendation for Einstein supporting him for membership in the Prussian Academy of Sciences, Planck and three other distinguished scientists Nernst, Rubens, and Warburg, discussed Einstein's contributions to relativity theory and cited the 'devastating results of [Einstein's] new interpretation of the time concept, which have repercussions on the whole of physics, and above all on mechanics and thence deep into epistemology . . .'.[3]

(*c*) *The Genesis of the Special Theory of Relativity.*

Zahar has once again raised the most interesting question as to how Einstein was led to propose the special theory of relativity. Within the context of Lakatos's and his own approach, the question is important because of the manner in which research programmes are judged. As Zahar has pointed out, because of his account of 'novelty', 'that *in order to assess the relation between theories and empirical data*

---

[1] Einstein [1905*b*], p. 38.

[2] Born [1969], p. 105.

[3] The letter is quoted in Seelig [1956], pp. 144–5.

within a research-programme, one has to take into account the way in which a theory is built and the problems it was designed to solve'.[1]

Zahar's account of the genesis of relativity theory cites two heuristical principles with which Einstein worked: (I) a maximisation of 'internal coherence', and (II) a trust only in symmetrical theories.[2] Zahar dwells on the asymmetries between the particle and the field, and between classical mechanics with its Newtonian relativity principle and the Maxwell–Lorentz electrodynamics with its special aether rest frame.

I am in full agreement with Zahar concerning the significance of these methodological criteria for Einstein. Nonetheless I differ with his account of the problems which were utmost in Einstein's thoughts while he was developing the special theory. I would like to present an alternative account which is built about the importance of Einstein's perception of the *ad hoc* character of the Lorentz theory in explaining the contraction hypothesis at both the macroscopic and microscopic (electron level) and on the relevance of the Michelson experiment in generating the paradox which forced Einstein to his new kinematics and time-concept. The account which I will thus present in the following pages is coherent with the theses proposed in the earlier sections of the present paper.

Before turning to the specific moves in the genesis of the special theory it will be useful to comment on some background considerations which appear to be relevant in the genesis, and which Zahar does not cite. These are both physical and epistemological in nature.

(*i*) Quantum theoretical considerations. It is clear that Einstein was strongly influenced by the electromagnetic picture of the world developing in the late nineteenth and early twentieth century, and Zahar notes this at several points. What he does not stress is what Einstein called the 'crisis in physics' caused by 'Planck's discovery of the quantum nature of energy'.[3] This discovery, Einstein saw more clearly than his contemporaries (and even Planck himself), raised very serious problems for the universal truth of both classical mechanics and electrodynamics. There is evidence that Einstein attempted to develop what he has termed a 'constructive' type of theory which would incorporate Planck's discovery as well as his own 1905 discovery of the quantum nature of light into a theory of stationary and moving body electrodynamics, both before his publication of the special theory of relativity and for several years after.[4] He was unsuccessful in these attempts and Einstein has noted that he had to move to a *different type* of theory, a 'principle-theory', in his inquiries in the electrodynamics of *moving* bodies. I shall comment on this type of theory more extensively below in my discussion of epistemological background to the genesis of the special theory of relativity.

(*ii*) Epistemological background. Einstein has noted (I shall quote him directly below) that the philosophical writings of Hume and Mach assisted him in his reanalysis of the concept of time. Einstein was never specific and precise in what he obtained from Hume and Mach, but there are sufficient suggestions in Einstein's writings that allow for the following reconstruction of Hume's and Mach's roles.

[1] Zahar [1973], p. 103.

[2] Zahar [1973], pp. 224–5.

[3] Einstein [1949], p. 37.

[4] See Wertheimer [1959], Ch. 10 and McCormmach [1970*b*].

In a letter to M. Besso written in 1948 he said:

> The extent to which . . . [Mach's *Mechanics* and *Wärmelehre*] influenced my own work is, to say the truth, not clear to me. As far as I am conscious of it, the immediate influence of Hume on me was greater.[1]

In his 'Autobiographical Notes' Einstein was somewhat more specific about what aspect of Hume's thought he found important:

> Hume saw clearly that certain concepts, as for example that of causality, cannot be deduced from the material of experience by logical methods.[2]

Though Mach's influence may well have been unconscious—it is difficult to determine this as there is very little that bears on this point written by Einstein in his early years—there is a methodological approach to the reanalysis of fundamental scientific concepts which one finds even in the early editions of Mach's *Science of Mechanics* which could not have failed to make some impression on Einstein, and which may well have served as an unconscious methodological guide for Einstein in his 1905 reanalysis of simultaneity.[3]

In his section on a 'Synoptical Critique of Newtonian Enunciations' Mach criticised Newton's mode of presenting his definitions and laws of mechanics, and suggested that even if the problems which he had raised earlier in the book regarding time and space were momentarily disregarded, that 'it is possible to *replace Newtonian enunciations by much more simple, methodologically better-arranged, and more satisfactory propositions*' (my italics).[4] What Mach then provided was a series of *definitions* which were sharply distinguished from *experimental propositions*. These sentences, Mach added, 'satisfy the requirements of simplicity and parsimony which, on economico-scientific grounds, must be exacted of them. They are, moreover, obvious and clear; for no doubt can exist with respect to any one of them either concerning its meaning or its source; and *we always know whether it asserts an experience or an arbitrary convention*'.[5]

Einstein would have found a similar approach taken in Mach's *Principien der Wärmelehre*, which he also read. Several quotations from the '*Wärmelehre*' will indicate the methodology which seems to have influenced Einstein in his analysis of the concept of time. In his chapter on a 'Critique of the Concept of Temperature' Mach stressed the 'conventional' and 'arbitrarily definitional' aspects of the concept and then added:

> It is remarkable how long a period elapsed before it definitively dawned upon inquirers that the designation of *thermal states* by *numbers* reposed on a *convention*. Thermal states exist in nature, but the concept of temperature exists only by virtue of our arbitrary *definition*, which could very well have taken another form. Yet until very recently inquirers in this field appear more or less unconsciously to have sought after a *natural* measure of temperature, a real temperature, a sort of Platonic Idea of temperature. . . .[6]

---

[1] Quoted by Holton [1968], p. 636.

[2] Einstein [1949], p. 13.

[3] Mach [1893].

[4] Mach [1893], p. 243.

[5] Mach [1893], p. 244.

[6] Mach [1896]. Quotations are from the English translation by T. J. McCormack [1903], p. 154.

Importantly, Mach also drew attention to parallels between the concept of temperature and the Newtonian kinematical concepts:

> Newton's conceptions of 'absolute time', 'absolute space', *etc.*, which I have discussed in another place, [in the *Science of Mechanics*] originated in quite a similar manner. In our conceptions of time the *sensation of duration* plays the same part with regard to the various measures of time as the sensation of heat played in the instance just adduced. The situation is similar with respect to our conceptions of space.[1]

The awareness that scientific theorising involves important definitional or conventional elements would have been underscored in Einstein's mind by his reading of Poincaré's *Science and Hypothesis* which we have evidence he became acquainted with, probably in 1903, as part of a discussion group.[2] Though we have no direct acknowledgment from Einstein of the influence of Poincaré on the methodological and epistemological background of his reasoning, it is interesting nonetheless to note that in the preface to *Science and Hypothesis* Poincaré wrote:

> We shall see that there are several kinds of hypotheses; that some are verifiable, and when once confirmed by experiment become truths of great fertility; that others may be useful to us in fixing our ideas; and finally that others are hypotheses only in appearance, and reduce to definitions or to conventions in disguise.[3]

Poincaré further added:

> In mechanics we shall be led to analogous conclusions, and we shall see that the principle of this science, although more directly based on experience, still share the conventional character of the geometrical postulates.[4]

It seems then that there were likely sources of conventionalist epistemology on Einstein in the years prior to his articulation of a new time concept, which, as we noted above, involves a conventionalist assumption in a most crucial way.

Another important epistemological facet of Einstein's background, and which he cites in connection with the genesis of the special theory of relativity, is the distinction between 'constructive' and formal or 'principle' theories in physics. Einstein asserts, and I shall quote him extensively below on this point in a more appropriate context, that it was only by proposing a 'universal formal principle' that he was able to resolve the problem of the electrodynamics of moving bodies. In 1919 in a note which he wrote for *The Times* of London Einstein amplified his views on these two approaches to physical theorising. Einstein wrote:

> We can distinguish various kinds of theories in physics. Most of them are constructive. They attempt to build up a picture of the more complex phenomena out of the materials of a relatively simple formal scheme from which they start out. Thus the kinetic theory of gases seeks to reduce mechanical, thermal and diffusional processes to movements of molecules—*i.e.*, to build them up out of the hypothesis of molecular motion. When we say that

---

[1] Mach [1896], p. 157.

[2] Poincaré [1902]. Page references are to the English translation *Science and Hypothesis*.

[3] Poincaré [1902], p. xxii.

[4] Poincaré [1902], p. xxvi.

> we have succeeded in understanding a group of natural processes, we invariably mean that a constructive theory has been found which covers the processes in question.
>
> Along with this most important class of theories there exists a second, which I will call 'principle-theories'. These employ the analytic, not the synthetic, method. The elements which form their basis and starting-point are not hypothetically constructed but empirically discovered ones, general characteristics of natural processes, principles that give rise to mathematically formulated criteria which the separate processes or the theoretical representations of them have to satisfy. Thus the science of thermodynamics seeks by analytical means to deduce necessary connections, which separate events have to satisfy, from the universally experienced fact that perpetual motion is impossible.
>
> The advantages of the constructive theory are completeness, adaptability and clearness, those of the principle theory are logical perfection and security of the foundations.
>
> The theory of relativity belongs to the latter class.[1]

Both 'principle' and 'constructive' types of theories are criticisable from two points of view. Einstein, in his 'Autobiographical Notes' proposed that empirical adequacy was most important in judging a theory: 'the theory must not contradict empirical facts'.[2] There is, however, ***another*** desideratum for adequate theories or as Einstein put it, another point of view from which one can judge a theory. This is the desideratum of 'inner perfection', also characterisable by such terms as 'naturalness' or 'logical simplicity' of the premises (of the basic concepts and of the relations between these which are taken as a basis).

With the introduction of these physical and epistemological facets of Einstein's background, I shall now turn to a reconstruction of the genesis of the special theory of relativity.

In his 'Autobiographical Notes' Einstein tells us in an often quoted passage that:

> Reflections of this type [and here Einstein is referring to problems which the quantum hypothesis had occasioned for Maxwell's and Lorentz's theories] made it clear to me as long ago as shortly after 1900, *i.e.*, shortly after Planck's trailblazing work, that neither mechanics not thermodynamics could (except in limiting cases) claim exact validity. By and by I despaired of the possibility of discovering the true laws by means of constructive efforts based on known facts. The longer and the more despairingly I tried, the more I came to the conviction that only the discovery of a universal formal principle could lead us to assured results. The example I saw before me was thermodynamics. The general principle was there given in the theorem: the laws of nature are such that it is impossible to construct a *perpetuum mobile* (of the first and second kind). How, then, could such a universal principle be found? After ten years of reflection such a principle resulted from a paradox upon which I had already hit at the age of sixteen: If I pursue a beam of light with the velocity c (velocity of light in a vacuum), I should observe such a beam of light as a spatially oscillatory electro-

---

[1] Einstein [1934], p. 73.

[2] Einstein [1949], p. 21.

> magnetic field at rest. However, there seems to be no such thing, whether on the basis of experience or according to Maxwell's equations. From the very beginning it appeared to me intuitively clear that, judged from the standpoint of such an observer, everything would have to happen according to the same laws as for an observer who, relative to the earth, was at rest. For how, otherwise, should the first observer know, *i.e.*, be able to determine, that he is in a state of fast uniform motion?
>
> One sees that in this paradox the germ of the special relativity theory is already contained. Today everyone knows, of course, that all attempts to clarify this paradox satisfactorily were condemned to failure as long as the axiom of the absolute character of time, *viz.*, of simultaneity, unrecognisedly was anchored in the unconscious. Clearly to recognise this axiom and its arbitrary character really implies already the solution of the problem. The type of critical reasoning which was required for the discovery of this central point was decisively furthered, in my case, especially by the reading of David Hume's and Ernst Mach's philosophical writings.[1]

This passage is unfortunately somewhat obscure and it has, as a result, given rise to a number of perplexities and misinterpretations. Since it does not mention specific experiments such as the Michelson–Morley experiment, and since it uses the phrase 'from the very beginning it appeared to me intuitively clear . . .', it has lent credence to anti-empiricist reconstructions of the genesis of relativity.[2] Let us recall what theoretical background Einstein had been exposed to and also what other accounts based on Einstein's own recollections tell us about this situation.

In 1916 Max Wertheimer had 'hours and hours' of conversations with Einstein about the genesis of the special theory of relativity which he recounted in a chapter in his later published book *Productive Thinking* (Wertheimer [1959]). In the same year Einstein completed a popular work on relativity *Über die spezielle und die allgemeine Relativitätstheorie*, in the preface of which Einstein stated that 'the author has spared himself no pains in his endeavour to present the main ideas in the simplest and most intelligible form, and on the whole, in the sequence and connection in which they actually originated'.[3] These accounts together with various other statements which Einstein made, and together with the evidence provided by the 1905 paper itself allows for the following type of reconstruction,

The light beam experiment which Einstein performed in his own mind at age 16 served to raise a *problem* but *not* to result in any *solution* at this point. Maxwell's equations *themselves* do not prohibit relative velocities of light less than or greater than c, in fact, taken together with the hypothesis of a stationary aether and Galilean transformations, they *entail* such velocities.[4] However, there is a principle of relativity for mechanical experiments performed in empty space which does prohibit experimental determinations of absolute rest: this is simply the proposition that Newton's laws of motion are invariant with respect to the Galilean transformations. Wertheimer questioned Einstein closely on this point:

---

[1] Einstein [1949], pp. 51–3. For further evidence of the effect which Planck's discovery and light quanta had on Einstein's approach to relativity see the reference cited in p. 75, footnote 4.

[2] For the most famous of such anti-empiricist analyses see M. Polanyi [1958], pp. 9–15. Also see Grünbaum's criticism in his [1961]. G. Holton seems to have some sympathy with anti-empiricist analyses of the genesis of special relativity in his fine article [1969]. For a discussion of the relative merits of Grünbaum and Holton on the genesis of special relativity see G. Gutting's [1972].

[3] Einstein [1917], p. *v*.

[4] This point is clearly made by Grünbaum [1961].

> When I asked him whether, during this [early] period, he had already had some idea of the constancy of light velocity, independent of the movement of the reference system, Einstein answered decidedly: "No it was just curiosity. That the velocity of light could differ depending on the movement of the observer was somehow characterised by doubt. Later developments increased that doubt."[1]

We may conjecture as to what those 'later developments' might have been. First, but not necessarily most importantly, it would seem that Lorentz's theorem of corresponding states for first order experiments associated with aberration would have corroborated Einstein's doubts. This theorem entails the apparent constancy of the velocity of light c to *first* order experiments of precision v/c. There are a number of experiments which, seen from the perspective of the Lorentz theory, indicate that the motion of the earth, and of matter in general, has no effect on light propagation or on the form of Maxwell's equations. To corroborate this influence a letter from Einstein to Shankland written in 1952 can be cited.[2] There Einstein noted that he 'was also guided [to relativity] by the result of the Fizeau experiment and the phenomenon of aberration'. Secondly, and perhaps more important, it is clear that Einstein was very forcefully struck by the electromagnetic induction experiments which indicated that only relative motion had any empirical effect on electromagnetic interactions. The received theories of Einstein's time, both Hertz's theory and more importantly Lorentz's theory, treated the situation of a magnet in motion and a wire at rest (in the aether) differently from a wire in motion and a magnet at rest.[3] Einstein was sufficiently impressed with the importance of this to begin his 1905 paper by citing such asymmetries which were not 'inherent in the phenomena'. In the same letter to Shankland quoted above, Einstein asserted: 'what led me directly to the Special Theory of Relativity was the conviction that the electromotive force induced in a body in motion was nothing else but an electric field.'[4] Finally, it is clear that Einstein *was* aware of the null result of the Michelson–Morley experiment. The general position which I believe can be extracted from the protracted discussion in the literature on this topic[5] is best put by saying that Einstein by the time he became aware of the experiment had already anticipated its null result. This is not particularly puzzling, however, as Poincaré too would have anticipated null results for the Trouton–Noble and Rayleigh–Brace experiments performed in the years 1902–4 on the basis of his own empirical generalisation of a principle of relativity.

Einstein's thinking which led to the reanalysis of simultaneity, then, was licensed by experimental considerations as enlightened by physical theory, and as further constrained by a desire for simplicity and a minimization of *ad hoc* or 'artificial' hypotheses. How such considerations might have led to the special theory of relativity and in particular the new analysis of the time concept will be taken up in the following pages.

---

[1] Wertheimer [1959], p. 215.

[2] Shankland [1964]. Also see the modified translation by Holton [1969], p. 156.

[3] See von Laue's analysis of this problem from the perspective of Hertz's theory in his [1911], p. 26. Also see the same problem dealt with from Lorentz's perspective in McCormmach [1970*b*], p. 56.

[4] See footnote 2.

[5] See the articles and references in p. 63, footnote 2 for an introduction to this issue, and of course Zahar [1973].

Wertheimer suggests that the Michelson–Morley experiment was no surprise to Einstein but that nevertheless it was 'very important and decisive'. This comment on the face of it does not appear to accord with a number of other comments which Einstein has made concerning the 'indirect' and 'negligible' effect which the Michelson–Morley experiment had on the genesis of relativity.[1] This apparent contradiction admits, I think, of a resolution which also clarifies the long quote from the 'Autobiographical Notes' above concerning the import of the light beam *gedankenexperiment* performed by Einstein at age 16.

We may hypothesise that on the basis of first order aether drift experiments and the electromagnetic induction analysis that Einstein had become convinced prior to learning of the Michelson experiment that light had no property of absolute rest. The Michelson experiment confirmed this, but in fact such a belief is stronger than anything which the Michelson experiment licenses, since the null result of the experiment can be explained by a contraction hypothesis. Such a route, however, would have been barred to Einstein by his reluctance to accept *ad hoc* hypotheses, and it may well be that this is why he did not move in the direction of developing a theory which would have employed both the contraction hypothesis and the second order 'local time' transformation. Such a 'doubly amended' aether theory[2] would have accounted for the fact that light *apparently* had no property of absolute rest, but it would be inconsistent with a principle of relativity interpreted so as to deny the existence of any special reference frame. That the principle of relativity, extraordinarily well confirmed for mechanical phenomena, was deeply entrenched in Einstein's mind is attested to by both Wertheimer's account and by Einstein's popular work of 1917. Thus in 1905 Einstein was confronted by the dilemma of interpreting how a light beam could have the same velocity for a reference system K′ as it did for a system K in motion with respect to K′. As above let us call this hypothesis of constant light velocity the *C.L.V.* hypothesis.

The *C.L.V.* hypothesis is, as mentioned above, confirmed by the Michelson or Michelson–Morley experiment, though it goes beyond it. The experiment, however, might offer a means of *concretising* the relations between the *C.L.V.* hypothesis and the assumptions of classical electromagnetic theory, as well as offering some empirical control over the speculations concerning light velocity in moving reference frames.

In the Einstein literature there seems to be only one source which explicitly touches on the crucial stage of Einstein's reasoning when he conjectured that the classical conceptions of time and simultaneity might require reanalysis. This source is Wertheimer's account.[3] Wertheimer couches his discussion of Einstein's reasoning in terms of an analysis of the Michelson experiment; a fact which is both interesting and troublesome in the light of Einstein's disavowal of the importance of the experiment. However, if we can assume that Einstein used the theoretical picture of the Michelson apparatus in connection with analysing the *C.L.V.* hypothesis, as essentially equivalent to his light beam *gedankenexperiment*, we reach a reasonable scenario which brings both the story told in the 'Autobiographical Notes' and the following Wertheimer account into accord. Wertheimer recounted Einstein's reasoning as follows (I have italicised certain passages in order to highlight the relation of the analysis to the *C.L.V.* hypothesis):

---

[1] See Shankland [1963] and Polanyi [1958], pp. 9–15, but also see Shankland [1973] for qualifications on his earlier notes.

[2] The term is Grünbaum's and appears in his [1964].

[3] Wertheimer [1959]. Quotations from Wertheimer given below are from pp. 218–19 of his book.

> Einstein said to himself: "Except for that [null] result, the whole situation in the Michelson experiment seems absolutely clear; all the factors involved and their interplay seem clear. But *are* they really clear? Do I really understand the structure of the whole situation, especially in relation to the crucial result?" During this time he was often depressed, sometimes in despair; but driven by the strongest vectors.
>
> In his passionate desire to understand or, better, to see whether the situation was really clear to him, he faced the essentials in the Michelson situation again and again, *especially the central point: the measurement of the speed of light under conditions of movement of the whole set in the crucial direction.* (my italics).
>
> This simply would not become clear. He felt a gap somewhere without being able to clarify it, or even to formulate it. He felt that the trouble went deeper than the contradiction between Michelson's actual and the expected result.
>
> He felt that a certain region in the structure of the whole situation was in reality not as clear to him as it should be, although it had hitherto been accepted without question by everyone, including himself. His proceeding was somewhat as follows: There is a time measurement while the crucial movement is taking place. "Do I see clearly," he asked himself, "the relation, the inner connection between the two, between the measurement of time and that of movement? Is it clear to me how the measurement of time works in such a situation?" *And for him this was not a problem with regard to the Michelson experiment only, but a problem in which more basic principles were at stake* (my italics).
>
> It occurred to Einstein that time measurement involves simultaneity. What of simultaneity in such a movement as this? To begin with, what of simultaneity of events in different places?
>
> He said to himself: "If two events occur in one place, I understand clearly what simultaneity means. For example, I see these two balls hit the identical goal at the same time. But . . . am I really clear about what simultaneity means when it refers to events in two different places? What does it mean to say that this event occurred in my room at the same time as another event in some distant place? Surely I can use the concept of simultaneity for different places in the same way as for one and the same place—but can I? Is it as clear to me in the former as it is in the latter case? . . . It is not!"

The above account is oversimplified, since it does not mention the problems with electromagenetic theory occasioned by Planck's quantum discovery and does not indicate the need to formulate a principle type of theory, and in addition does not cite the Fizeau experiment or the electromagnetic induction analysis.

There is, however, no reason to demand that it do so as the account is concerned with a very specific point in Einstein's reasoning and accords with our other information which can be understood as playing an earlier and background role at this point.

The scenario is conjectural but has several interesting features. First, it does give us a plausible reading of the available documents, fills in some gaps, and resolves some paradoxes. Secondly, it suggests that there are some interesting parallels between the discovery of the special theory of relativity and the way in which Einstein presented the theory in its kinematical aspect in his 1905 paper. Finally it may not be unimportant that, though the above scenario was developed prior to the availability of Shankland's very recent and significant 'Conversations with Albert Einstein, II', it would appear that this new evidence is in complete accord with the above reconstruction.[1] Shankland finds the Wertheimer account completely reliable, and further adds that Einstein had told him at the crucial point in his reasoning that ' "at last it came to me that [absolute] time was suspect," and that the new absolute must be the speed of light in vacuum.'[2]

2 *The Methodology of Comparing Competing Theories.*

2.1 *The Concept of an* Ad Hoc *Hypothesis.*

Earlier, in connection with my account of Lorentz's and Einstein's theories, I introduced Zahar's tripartite analysis of *ad hocness*. In the context of that discussion I argued that (*i*) even on *his* analysis, Lorentz's programme after 1904 involved an *ad hoc* assumption of the generalised *M.F.H.*, and (*ii*) Einstein's light postulate, contrary to Zahar's interpretation, possessed considerable independent support. I would now like to reconsider Zahar's analysis from a more general methodological perspective.

(*a*) Ad hocness *is an attribute of an hypothesis.*

The first suggestion I would make is to attribute the characteristic of *ad hocness* to a hypothesis—a constituent element of a theory—and only derivatively to a theory. As noted above, Zahar applies the term explicitly only to theories within research programmes, and only at most derivatively to hypotheses. My reasons for making this suggestion are the following.

First, consider a theory T_1 which is confronted by a *prima facie* falsifying result. This result may be either experimental *or theoretical*—*e.g.*, a result derived from *another* well corroborated theory possibly in a different subject area. The proponent of T_1 often modifies T_1 to a T_1' by altering one or more of the constituent elements—*i.e.*, of the *hypotheses*. Other scientists will scrutinise these constituents of T_1' to determine if the modifications possess independent support (say from another theory) or if they provide additional interesting and/or testable results. If the answer is positive, T_1' is accepted as an advance; if the answer is negative, the *modification*—the hew hypothesis—is termed *ad hoc*. On the basis of scientific behaviour then, *ad hocness* appears to be a property of specific hypotheses, as embedded, however, in a constellation of other hypotheses constituting the theory.

---

[1] Shankland [1973].

[2] Shankland [1973], p. 897.

The second reason for suggesting that *ad hocness* be ascribed to specific hypotheses depends on an analysis of hypothesis and theory confirmation which will be discussed in detail below in section 2.2. Suffice it for now to say that from the perspective of the Bayesian approach suggested there, it is crucial to be able to segregate a new hypothesis from a theory and to assess the *hypothesis'* empirical consequences and its 'prior probability' which is a function of its logical simplicity and its intertheoretical relations. I shall argue below that it is only by this type of approach that *ad hoc* hypotheses can be judged in a *quantitative* as well as a qualitative manner, and the *seriousness* of the *ad hoc* character of a hypothesis, and derivitively of the theory in which it is embedded, is analysable.

(*b*) *What makes an hypothesis* ad hoc?

In the section immediately above when I described the behaviour of scientists in ascribing *ad hocness* to a hypothesis, I suggested that an hypothesis is *ad hoc* if it is proposed as a modification of a theory faced with a *prima facie* falsifying result of either experimental or intertheoretical origin, and if the new hypothesis does not possess independent theoretical or experimental support. Let me elaborate on these aspects of *ad hocness* and relate them to Zahar's three senses of *ad hoc*.

First, it would seem that a judgment of *ad hocness* cannot be an objective characteristic but rather must be *intersubjective*. If a hypothesis H of theory T is modified to H' under the conditions just discussed, and if no one in the scientific community can discern independent support or tests for H', then H' is *ad hoc*. If later, an argument is constructed for H' which provides it with independent theoretical support or if a new test is envisaged, the hypothesis H' loses its *ad hoc* character. It does not appear to be possible at this point in the development of the logic of science to provide an account of *ad hocness* which will possess greater foresight about hypotheses than do the best minds of a scientific community. Further, if one wishes to show that the rule to avoid *ad hoc* hypotheses functioned *historically* to rationalise specific moves by scientists, such as Einstein, then one must accept the judgments of those scientists made at that historical time. This indicates that Zahar's first sense of *ad hoc*, in which a theory is *ad hoc* if 'it has no novel consequences as compared with its predecessor', is a logical dream. (Zahar illustrates this sense by arguing, as Grünbaum did earlier,[1] that the Kennedy–Thorndike interferometer experiment was an objectively independent test of the *L.F.C.* even though it had not occurred to scientists in the years when Lorentz's and Einstein's theories were in strong competition.)

Let me now consider Zahar's third sense of *ad hoc* before turning to his second sense.

As quoted earlier, Zahar believes that a theory is *ad hoc*$_3$ 'if it is obtained from its predecessor through a modification of the auxiliary hypotheses which does not accord with the spirit of the heuristic of the programme'. It seems to me that this is too vague, in fact vague to the point of inapplicability. It may be that the concept, 'the spirit of the heuristic', could be made more precise, but as such it is difficult to determine what it permits and what it excludes. Consider, for example, the heuristic of the 'ether programme' as characterised by Zahar, *viz*. that 'requires a physical phenomena be explained in terms of actions propagated in the ether'. This would seemingly exclude action-at-a-distance forces, but such forces functioned as an integral part of Helmholtz's theory of electrodynamics which incorporated both fields (or ethereally transmitted forces) and action-at-a-distance forces in one theory. Both Hertz and Lorentz thought highly of this theory throughout the 1870s and into the mid 1880s, and came to appreciate Maxwell's theory—which is a limiting case of Helmholtz's theory ([1870])—through their many years work with Helmholtz's theory.[1] Neither scientist, to the best of my knowledge, criticised Helmholtz for incorporating an *ad hoc* assumption about action-at-a-distance into his theory.

---

[1] Grünbaum [1959] and [1963].

Let us now examine Zahar's *ad* hoc_2 which ascribes *ad hocness* to a theory 'if none of a new theory's novel predictions have been actually verified' (or worse, if they have been falsified). I think that this sense of *ad hoc* is close to the sense in which *ad hoc* is used in science, but that Zahar's analysis is weakened by his account of novelty and by what I will term an 'unicriterial tunnel vision', which I believe to be the consequence of an unconscious positivism.[2]

I believe that 'novelty', in connection with 'novel facts' or facts which offer empirical support to a theory or hypothesis, is best examined within a broader context of confirmation or corroboration in general.[3] Accordingly my analysis of *ad* hoc_2 will require the development of some notions of what I have previously termed a 'logic of comparative theory evaluation', and also a discussion of the formal scheme for theory corroboration.

2.2 *The Formalism of a Logic of Theory Competition.*

I should like to begin by stating a formal theorem of the probability calculus known as the *general division theorem.* This theorem, which in an expanded form is known as Bayes's theorem for inverse probability, has often been thought to capture important aspects of the concept of hypothesis confirmation. I am going to use a formalism which is quite close to Carnap's [1950] and [1962] notation,[4] but which employs different letters so as to facilitate the interpretation I wish to defend, and which owes something to Popper's brief remarks (in his [1962]) on a quantitative measure of the severity of an empirical test of a hypothesis.[5] The general spirit of the approach which I shall recommend is in basic agreement with Salmon's eloquent defense of the utility of Bayes's theorem to elucidate the various aspects of confirmation and the hypothetico-deductive method of scientific inquiry.[6] I believe that I can show that the interpretation of the theorem

---

[1] See Hirosige [1969] and McCormmach [1970*a*] for discussions of the early work of Lorentz and his commitment to Helmholtz's theory.

[2] In several papers, Schaffner [1970], [1973], [1974*b*], I have argued for the importance of such trans-empirical criteria as simplicity and intertheoretic accord in both the genesis and justification of scientific theories. I believe that a stress on only one—the empirical—aspect of constraints on theories blinds one to the way theories are in fact developed and accepted in science. 'Unicriterial tunnel vision' is a phrase which captures this focus on only one criterion of theory genesis and/or choice.

[3] I do not think that the position to be taken in the following pages requires a choice, at this time, between a sophisticated inductivist or confirmationist approach, and a falsificationist approach to theory testing. I use the terms confirmation *or* corroboration since the interpretations of the expressions in the general division theorem developed *below* seem neutral to this distinction.

[4] Cp. Carnap [1962], pp. 326–37.

[5] Popper [1962], pp. 390–1.

[6] Salmon [1967], esp. pp. 115–31.

I will provide is stronger than Salmon's, and that it clarifies the notion of novelty, of an *ad hoc* hypothesis, of trans-empirical constraints on a scientific theory, and of the reason why Einstein's theory replaced Lorentz's theory.

I shall state the general division theorem without proof inasmuch as it is a noncontroversial theorem in the probability calculus. Let us use the letter T for any theory in question and divide it into two parts: a hypothesis H and the remaining hypotheses, represented by the letter θ. We shall also write T' for a later theory obtained from T by substituting a new hypothesis H' in lieu of H but keeping θ the same. The letter b will represent our ***background knowledge*** and the letter e will stand for the positive outcome of some experiment ***not*** part of b. The background knowledge will contain (*1*) a set of theories ***other than*** the theories which are being subjected to evaluation, (*2*) descriptions of relevant experiments (though not of e in accordance with the earlier stipulation), and (*3*) a list of criteria which a good theory ought to possess and in terms of which it is possible to estimate the plausibility of a theory T prior to an explicit consideration of the effect of e on the theory. I shall discuss these criteria below. The general division theorem using these notions can then be written as:

$$p(T/b \,\&\, e) = \frac{p(T/b)\,.\,p(e/T \,\&\, b)}{p(e/b)}$$

The p or probability will here be taken to represent the degree of belief in the truth of the numerator conditional on the truth of the denominator. It would take me beyond the scope of this paper to consider the various interpretations of probability in association with the position to be taken here, and it seems sufficient to accept a personalist interpretation with the proviso that b contains a list of intersubjectively agreed upon constraints which will cause the convergence of the degrees of belief among scientists.[1]

Several comments about each of the terms in the general division theorem are needed. The expression $p(T/b \,\&\, e)$ represents the probability of T after e has been taken into account and in the light of background knowledge b. The term $p(T/b)$ reflects the 'prior' assessment of T in the light of other theories, experiments, and criteria of theory evaluation. The term $p(e/T \,\&\, b)$ is generally called the 'likelihood' of obtaining e on the assumption of $T \,\&\, b$; it can be set equal to or very close to 1, if T explains e. The expression $p(e/b)$ has been termed the 'expectedness' and represents the probability of obtaining e on the basis of theories ***other than*** T (and of other experiments and criteria of theory evaluation).[2] I should now like to argue that this interpretation of the general division theorem embodies a concept of novelty and can be used to analyse the notion of an *ad hoc* hypothesis in both qualitative and quantitative aspects.[3]

---

[1] On this interpretation b contains the corpus of well corroborated theories which exercise intertheoretical constraints on T, θ, H, and H', as well as general principles for determining relative simplicities of competing theories, as well as previous experimental data.

[2] If more explicit detail is desired, the expression $p(e/b)$ can be expanded, as it is in the usual form of Bayes's theorem, to explicitly reflect the specific n alternative competing theories, $T_1, T_2, \ldots T_n$, as:

$$p(e/b) = \sum_{i=1}^{n} p(T_i/b)\,.\,p(e/T_i \,\&\, b).$$

[3] My interpretation of the expressions in the general division theorem differs from Salmon's [1967], which is closest, in eliminating his truth predicate (Salmon [1967],

The concept of a novel or theory supporting fact e is represented in the general division theorem by a low value of $p(e/b)$ and a high value of $p(e/T$ and $b)$. It is interesting to note that this is precisely Popper's ([1962]) requirement for a severe test of a hypothesis, and it is to be expected that a novel fact is one which tests a theory more severely and which, if the theory explains it, corroborates the theory to a high degree.

A hypothesis H' is *ad hoc* from the perspective of this analysis if it is the consequence of the *prima facie* falsification of theory T, such that it is proposed as a modification of T, and there is no additional support for H' in b, either of a theoretical or experimental nature. This characterisation of *ad hocness* represents the *historical* aspect of an *ad hoc* hypothesis. It should be noted, however, that there is a *logical* aspect to *ad hocness* which is importantly *a*temporal, and which does not depend on whether H' was proposed after the falsification of T or before either e became available or came under explicit consideration. This logical aspect of *ad hocness* is contained in the formalism of the general division theorem. The 'prior' probability of T' is necessarily equal to the prior probability of the conjunction of θ and H'. H' is logically independent of θ, since if it were not, either (*1*) H' would be derivable from θ and would not then be *ad hoc*, or (*2*) H' would contradict θ, and would not then be conjoinable with θ so as to allow $\theta \,\&\, H'$ to entail e, except trivially. The rule for the calculation of conjunctions of independent probabilities, as applied to this case and using conditional probabilities, allows us to write:[1]

$$p(T'/e \,\&\, b) = \frac{p(\theta \,\&\, H'/b)\,.\,p(e/T' \,\&\, b)}{p(e/b)} = \frac{p(\theta/b)\,.\,p(H'/b)\,.\,p(e/T' \,\&\, b)}{p(e/b)}.$$

But $p(H'/b)$ is necessarily close to zero, since by assumption, e is the only empirical evidence for H' (and H' possesses no theoretical support).[2] Therefore the expression on the right hand side of the equation becomes close to zero, and, *a fortiori*, e does not confirm T'.[3] Clearly it is desirable that this be the case since

---

p. 117) and *explicitly* introducing the background knowledge into the formalism. This is crucial for the line of reasoning presented below. (See my [1968] review of (Salmon's [1967] for an earlier criticism of his interpretation of Bayes's theorem.)

[1] If it is felt that the assertion of H' (as in the expression $p(H'/b)$) *without* the assumption of θ will not represent a fair test of H', since a number of a theory's hypotheses are usually employed in deducing consequences, it should be noted that the more general rule for *non*-independent conjunctions is:

$$p(H' \,\&\, \theta) = p(\theta)\,.\,p(H'/\theta), \text{ or assuming } b:$$
$$p(H' \,\&\, \theta/b) = p(\theta/b)\,.\,p(H'/\theta \,\&\, b).$$

[2] In my approach to the logic of comparative theory evaluation using the general division theorem, it should be clear that I am neither providing a fully formally characterised language nor a measure function which would assign prior probabilities to hypotheses in anything like a Carnapian sense. Rather I am using the formalism of the probability calculus as a unifying and precision producing heuristic guide to this 'logic'. For example, the probability $p(H'/b) \approx 0$ does not follow from the logical universality of H' and the non-universal character of b, but follows rather from the consensus of informed scientists that H' possesses essentially *no* support except for e.

[3] This argument assumes that $p(e/b)$ is significantly larger that $p(H'/b)$. In the historical case we will consider this was in fact the situation, as on Stokes's theory of aether drag, for example, one would expect a null result (e) for the Michelson-Morley experiment. The reason why Stokes's theory is to be considered here is given by the expression for

e could only confirm T' by virtue of its incorporation of H', which is *ad hoc*. Note that according to the logical formalism H' need not have been elicited by a clash between e and T; H' could have appeared earlier in the scientific literature. If, however, $p(H'/b)$ is at any and all times close to zero, T' cannot be confirmed by e. The fact that H was historically elicited by e falsifying T does make H' historically *ad hoc*, and this temporal property of H' probably does increase its lack of attraction for scientists. This would seem to be more psychological than logical, but still important from a historical point of view.

According to the formalism developed e cannot confirm T' because e is the only experimental consequences of H', and it is excluded by the form of the general division theorem here used from appearing in the background knowledge b. It is historically incontrovertable though that such a T' can be further confirmed by additional experiments. Let us examine how the formalism reflects this.

To determine the effect of a new experiment e' on T', we form a new background knowledge b', now *including* e. The general division theorem then is written as:

$$p(T'/e' \,\&\, b') = \frac{p(\theta/b') \,.\, p(H'/b') \,.\, p(e'/T' \,\&\, b')}{p(e'/b')}$$

using the expanded form of T' introduced above, *mutatis mutandis*. Now the term $p(H'/b')$ is somewhat larger than $p(H'/b)$ (approximately by a factor $(p(e/b))^{-1}$) since $b' = b\&e$, the increase being given by Bayes's theorem. The value of $p(H'/b')$ is still low however, and thus H' still continues to weaken T'. We can however consider T' *vis à vis* its rivals in connection with the new experiment e' by examining the "odds form" or ratio of $p(T'/b'\&e')/p(T_i/b'\&e')$, where $T_i \neq T'$. This ratio equals $p(H'/b').p(\theta/b').p(e'/b'\&T')/p(T_i/b').p(e'/T_i \&b')$. If T' is the only theory conveying a high likelihood on e', then the likelihood expression can "swamp" the weakening effect of H' on T' and result in a high probability of T' relative to its rivals. This abstract situation was actually realised *in concreto* when Lorentz's theory with its *L.F.C.* was highly confirmed by being able to explain the Zeeman effect. (In this case $e =$ the Michelson experiment, $H' =$ the *L.F.C.*, and $e' =$ the Zeeman effect.)

The hypothesis H', even though it allows T' to maintain some acceptability by screening off the falsifying effect of e on its earlier form T, is likely to be considered suspect and as weakening the theory T'. It is of interest to note that the analysis so far provided permits an account of the seriousness of the *ad hoc* character of such a hypothesis. If H' is *ad hoc*, however, an assessment of its demerits cannot be based on empirical considerations, since by supposition, e was the only experimental result associated with H'. It is at such a point that what I have termed trans-empirical considerations enter into the asessment of an *ad hoc* hypothesis, and allow, I believe, a determination of the seriousness of the *ad hoc* character of H'.

There are two types of considerations which are different from direct empirical

---

expansion of $p(e/b)$ developed in footnote 2, page 00 above. It should be added here that the sense in which e does *not confirm* T' is to be taken as relative to the prior *degree* of confirmation which T possessed relative to b, *i.e.*, before the appearance of e. Thus e does not confirm T' since $p(T'/b\&e) << p(T/b)$.

support which significantly affect a theory or hypothesis. These represent (*i*) the effect which other well confirmed (or corroborated) theories have on a particular theory in question, and (*ii*) the simplicity which a theory possesses relative to its competitors. It would take me too far afield to comment on these considerations other than in the manner in which they specifically affected the acceptance of Einstein's theory, and this I shall do below in section 3. Suffice it to say that the new hypothesis H' introduced above will be judged as more or less acceptable depending on (*i*) the theoretical accord or theoretical discord which it introduces, and on (*ii*) the simplicity of its entities and logical form. If intertheoretic accord/discord and simplicity can be quantified, as I believe they can, then we possess a means of judging the seriousness of an *ad hoc* hypothesis. This becomes particularly important if all competing theories which are attempting to explain a set of experiments are thought to be *ad hoc* in certain respects.

3 '*Why Did Einstein's Programme Supersede Lorentz's?*'

Zahar has suggested that the reason why Einstein's programme replaced Lorentz's is due to the 'greater heuristic power' of Einstein's programme, which he conceives of as "*based on two distinct requirements, (1) a new law should be Lorentz-covariant and (2) it should yield some classical law as a limiting case*'.[1] Zahar also contends that though heuristic power is important, the 'ultimate' judgment of a research programme's superiority must be empirical, and that 'Einstein's relativity programme superseded Lorentz's in the empirical sense in 1915 with its explanation of the precession of Mercury's perihelion'.[2]

I would like to argue though Einstein's special theory of relativity does present, quite specifically, a method for working out problems in the electrodynamics of moving bodies,[3] and that this can be termed a heuristic, that the reasons for Einstein's approach replacing Lorentz's are not sufficiently captured by Zahar's suggestions. As I have noted several times above, many scientists became convinced of the superiority of Einstein's theory well before 1915, and they did so on what I termed in section 2 *trans-empirical considerations*. Let us examine the situation in the light of the historical and methodological analyses presented above.

3.1. *The Simplicity of Einstein's Theory.*

In my view, one of the fundamental reasons why Einstein's theory attracted the attention and commitment of early twentieth-century physicists was because of its simplicity in comparison with Lorentz's theory. Zahar treats of simplicity as derivative:

---

[1] Zahar [1973], p. 243.

[2] Zahar [1973], p. 249.

[3] Einstein, in his [1905*b*], p. 59, after he has worked out the pressure of radiation on perfect reflectors asserts:

> All the problems in the optics of moving bodies can be solved by the method here employed. What is essential is, that the electric and magnetic force of the light which is influenced by a moving body, be transformed into a system of coordinates at rest relatively to the body. By this means all problems in the optics of moving bodies will be reduced to a series of problems in the optics of stationary bodies.

This is possibly the clearest statement of a 'heuristic' in Einstein's writings. It is, however, difficult to see that Poincaré, proceeding from a Lorentzian point of view, would not employ an identical heuristic.

> According to positivists like Mach, the mere elimination of . . . [(physically) uninterpreted mathematical entities] increases simplicity, and thereby constitutes progress. My claim is that such eliminations are *by-products* of new research programmes whose heuristic eliminates certain entities. This may be accompanied by a—contingent—increase in simplicity.[1]

It seems to me that this type of simplicity which Zahar cites is only one aspect of the simplicity which Einstein's theory possessed. Let me suggest what I take the simplicity to involve.

First, by beginning with his two well supported postulates—the relativity principle and the light postulate—Einstein commenced with a small number of hypotheses. By his bold reanalysis of the concept of simultaneity he was able to show that there were no ontologically privileged times or length measurements associated with a special reference frame. This eliminated the aether in the classical Lorentz sense and also introduced symmetry into the time and space transformations. The specific form of the transformations were derived from these considerations and a hypothesis of the isotropy of space. The principle of relativity was applied to Maxwell's and to Lorentz's equations for stationary bodies and with the new kinematics *generated* the electric and magnetic force transformations which Lorentz had had to postulate; the same considerations generated the charge density transformation. Einstein, in the last section of his paper showed that the Lorentz force was derivable, and need not be postulated independently. Also the Einstein theory allowed, in Minkowski's hands, for a *unification* of the laws of the conservation of energy and of momentum. The relativity principle, in conjunction with the laws of conservation of energy and momentum, entails the equivalence of mass and any form of energy, or the principle of 'the inertia of energy'. As Born has noted, 'the formula [$E = mc^2$] had been known for special cases 'prior to Einstein's theory,[2] but the generality of Einstein's and other relativity theoriests' proofs was of considerable significance, and Pauli has termed it 'the most important of the results of the theory of special relativity'.[3] The point of these remarks is to show the *systematicity*—the simplicity of system—which Einstein's theory introduced into physics in the first few years after its promulgation. M. von Laue, in his ([1911]) textbook—the first on relativity theory—wrote:

> Though a true experimental decision between the theory of Lorentz and the theory of relativity is indeed not to be gained, and that the former, in spite of this, has receded into the background, is chiefly due to the fact that, close as it comes to the theory of relativity, it still lacks the great simple universal principle, the possession of which lends the theory of relativity . . . an imposing appearance.[4]

Einstein's theory has to be contrasted with Lorentz's in this regard. Lorentz's theory involved considerable less simplicity of system and at least one *ad hoc* hypothesis, the generalised *M.F.H.* of 1904, which was seriously *ad hoc* in the context of his theory. To show that this was the case let us turn to consider the effect of the theoretical context in 1910–12.

[1] Zahar [1973], p. 239.

[2] Born [1969], p. 109.

[3] Pauli [1921], p. 123.

[4] von Laue [1911], pp. 19–20.

3.2. *Intertheoretical Considerations Favouring Einstein over Lorentz.*

The development of quantum mechanics in the early twentieth century began to cast serious doubt on the general foundations of the Lorentz approach. The Maxwell theory, and Lorentz's theory as well, yields the incorrect Rayleigh–Jeans law for black-body emission. Einstein had shown, in a controversial paper in 1905,[1] that there were definite limits to the classical electromagnetic theories with his introduction of light quanta. Planck was preoccupied with the relationship between the electron theory and quantum theory and tried three successive times from 1910 to 1914 to produce a coherent theory. In connection with his 'second theory' of 1911 Jammer notes that

> he proposed to derive the radiation law in a way which he thought is 'free of inner inconsistencies' and which 'does not depart from the core of the classical electrodynamics and electron theory more than is absolutely necessary in view of the irreconcilable differences with the quantum hypothesis'.[2]

In 1911 at the first Solvay Congress in Brussels, a number of eminent physicists convened to discuss the quantum theory. Lorentz and Einstein attended, and in his opening remarks Lorentz drew attention to the problems associated with black-body radiation. Several weeks after the Congress had concluded Poincaré published a paper in which he demonstrated that the hypothesis of light quanta is essential if one accepts Planck's radiation law. It was becoming increasingly evident that Lorentz's constructive approach to nature was seriously deficient inasmuch as it was inconsistent with quantum theory.

If we examine the changing background assumptions of this time I believe that we should see the weak rationale for some of Lorentz's hypotheses becoming even weaker. Consider the generalised *M.F.H.* discussed earlier. This hypothesis is *ad hoc* but it *might* be said to possess a slight degree of theoretical support on the supposition that Poincaré's arguments concerning the non-electromagnetic nature of the force holding the electron together were misguided. If all forces were electromagnetic the generalisation of the *M.F.H.* would not be *ad hoc* but might follow from a deeper classical theory. The calling into serious question of the fundamental assumptions of the Lorentz electron theory by light quanta made such *ad hoc* assumptions seriously *ad hoc*, by introducing in advance, counterarguments to any such deeper classical theory.

Einstein's theory of special relativity on the other hand was not troubled by the existence of light quanta. Though it led to the same predictions as Lorentz's theory concerning the motion of the electron, it did so, Einstein wrote 'without requiring any special hypothesis whatsoever as to the sturcture and the behaviour of the electron.[3] Since it was not a 'constructive theory of electrodynamics', but was rather a more general 'principle-type' of theory, in reality concerned with space and time, it could be applied to the *approximately true* theories of Maxwell and Lorentz without becoming encumbered by their deficiencies in the quantum arena.[4] By 1910 it had become clear that Einstein's theory could be applied not only in electrodynamics, but also *directly* in mechanics, as it was by Lewis and Tolman,[1] and in thermodynamics by Planck.[2] These changes in the theoretical context or background theoretical knowledge seemed to make Einstein's theory much more acceptable than the more limited and by now more suspect Lorentz theory.

---

[1] Einstein [1905*a*]. [2] Jammer [1966], p. 47. [3] Einstein [1916], p. 51.

[4] The quantum limitations of Maxwell's and Lorentz's theories were realised by Einstein very early. See his comments in his [1949], pp. 37–55 and his letter to Seelig in 1955, quoted in Born [1969], p. 104.

KENNETH F. SCHAFFNER
University of Pittsburgh

REFERENCES

Agassi, J. [1964]: 'The Nature of Scientific Problems and their Roots in Metaphysics', in M. Bunge (*ed.*): *The Critical Approach to Science and Philosophy*, pp. 189–211.

Born, M. [1969]: *Physics in My Generation.*

Brace, R. B. [1904]: 'On Double Refraction in Matter Moving Through the Aether', *Philosophical Magazine*, **6**, pp. 317–29.

Bucherer, A. H. [1908]: 'Messungen an Becquerel strahlen. Die experimentelle Bestätigung der Lorentz—Einsteinschen Theorie', *Physikalischer Zeitschrift*, **9**, pp. 755–62.

Carnap, R. [1962]: *Logical Foundations of Probability.*

Einstein, A. [1905*a*]: 'Über eine die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtpunkt', *Annalen der Physik*, **17**, pp. 132–48.

Einstein, A. [1905*b*]: 'Zur Elektrodynamik bewegter Körper', *Annalen der Physik*, **17**, pp. 891–921. Page references are to the English translation in A. Einstein *et al.*, *The Principle of Relativity*, pp. 35–65.

Einstein, A. [1905*c*]: 'Ist die Trägheit eines Körpers von seinem Energiegehalt abhängig?' *Annalen der Physik*, **18**, pp. 639–41.

Einstein, A. [1907]: 'Relativitätsprinzip und die aus demselben gezogenen Folgerungen', *Jahrbuch der Radioaktivität*, **4**, pp. 411–62.

Einstein, A. [1917]: *Über die Spezielle und die Allgemeine Relativitätstheorie, Gemeinverständlich.* Page references are to the English translation *Relativity: the Special and General Theory.*

Einstein, A. [1934]: *The World as I See It.*

Einstein, A. [1949]: 'Autobiographical Notes', in P. A. Schilpp (*ed.*): *Albert Einstein: Philosopher-Scientist*, pp. 1–95.

Einstein, A. [1950]: *Out of My Later Years.*

Feyerabend, P. K. [1964]: 'Realism and Instrumentalism: Comments on the Logic of Factual Support', in M. Bunge (*ed.*): *The Critical Approach to Science and Philosophy*, pp. 280–308.

Grünbaum, A. [1959]: 'The Falsifiability of the Lorentz–Fitzgerald Contraction Hypothesis', *The British Journal for the Philosophy of Science*, **10**, p. 48.

Grünbaum, A. [1961]: 'The Genesis of the Special Theory of Relativity', in H. Feigl and G. Maxwell (*eds.*): *Current Issues in the Philosophy of Science*, pp. 43–53.

Grünbaum, A. [1963]: *Philosophical Problems of Space and Time.*

Grünbaum, A. [1964]: 'The Bearing of Philosophy on the History of Science', *Science*, **143**, pp. 1406–12.

Gutting, G. [1972]: 'Einstein's Discovery of Special Relativity', *Philosophy of Science*, **39**, pp. 51–67.

Helmholtz, H. von [1870]: 'Über die Bewegungsgleichungen der Elektricität für ruhende leitende Körper', *Borchardts Journal für die reine und angewandte Mathematik*, **72**, pp. 57–129.

Hirosige, T. [1969]: 'Origins of Lorentz' Theory of Electrons and The Concept of the Electromagnetic Field'. *Historical Studies in the Physical Sciences* **1**, pp. 151–210.

Holton, G. [1968]: 'Mach, Einstein, and The Search for Reality', *Daedalus* **97**, pp. 636–73.

Holton, G. [1969]: 'Einstein, Michelson, and the "Crucial" Experiment'. *Isis*, **60**, pp. 133–97.

---

[1] Lewis and Tolman [1909].

[2] Planck [1907].

Holton, G. [1973]: *Thematic Origins of Scientific Thought: Kepler to Einstein.*

Jammer, M. [1966]: *The Conceptual Development of Quantum Mechanics.*

Keswani, G. H. [1965]: Origin and Concept of Relativity', *The British Journal for the Philosophy of Science*, **15**, pp. 286–306, and **16**, pp. 19–32 and pp. 273–94.

Kuhn, T. S. [1962]: *The Structure of Scientific Revolutions.*

Lakatos, I. [1968]: 'Criticism and the Methodology of the Scientific Research Programmes', *Proceedings of the Aristotelian Society*, **69**, pp. 149–86.

Lakatos, I. [1970]: 'Falsification and the Methodology of Scientific Research Programmes', in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge*, pp. 91–195.

Lewis, G. N. and Tolman, R. C. [1909]: 'The Principle of Relativity and Non-Newtonian Mechanics', *Philosophical Magazine*, **18**, pp. 510–23.

Lorentz, H. A. [1892*a*]: 'La Théorie Electromagnétique de Maxwell et son Application aux Corps Mouvants', *Arch Neerl.* **25**, pp. 363–553.

Lorentz, H. A. [1892*b*]: 'The Relative Motion of the Earth and the Ether', *Versl. Kon. Akad. Wetensch. Amsterdam* **1**, pp. 74–8.

Lorentz, H. A. [1895]: '*Versuch einer Theorie der Elektrischen und Optischen Erscheinungen in Bewegten Körpern.*

Lorentz, H. A. [1899]: 'Simplified Theory of Electrical and Optical Phenomena in Moving Systems', *Koninkl. Akad. Wetenschap. Proc.* (*English edition*), **1**, pp. 427–400. Reprinted in Schaffner [1972], pp. 255–273.

Lorentz, H. A. [1904]: 'Electromagnetic Phenomena in a System moving with any velocity smaller than that of light', *Proceedings of the Royal Academy of Amsterdam*, **6**, pp. 809–34. Page references are to the version reprinted in A. Einstein *et al*, *The Principle of Relativity.*

Lorentz, H. A. [1909]: *The Theory of Electrons.*

Lorentz, H. A. [1914]: *Das Relativitätsprinzip, Drei Haarlemer Vortrage.*

Lorentz, H. A. [1915]: *The Theory of Electrons*, (second edition).

McCormmach, R. [1970*a*]: H. A. Lorentz and the Electromagnetic View of Nature', *Isis*, **61**, pp. 459–97.

McCormmach, R. [1970*b*]: 'Einstein, Lorentz, and the Electron Theory', *Historical Studies in the Physical Sciences*, **2**, pp. 41–87.

Mach, E. [1893]: *The Science of Mechanics*, second edition.

Mach, E. [1896]: *Principen der Wärmelehre.*

Mermin, N. D. [1968]: *Space and Time in Special Relativity.*

Minkowski, H. [1908*a*]: 'Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern', *Göttingen Nachrichten*, pp. 53–111.

Minkowski, H. [1908*b*]: 'Raum und Zeit'. Translated version (in English) appears in A. Einstein *et al.*, *The Principle of Relativity*, pp. 75–91.

Møller, C. [1952]: *The Theory of Relativity.*

Pauli, W. [1921]: 'Relativitätstheorie', *Encyklopädie der mathematischen Wissenschaften*, **19**, pp. 543–775. Page references are to the English translation, [1958]: *Theory of Relativity.*

Planck, M. [1907]: 'Zur Dynamik Bewegter Systeme,' *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, **29**, pp. 542–70.

Poincaré, H. [1900]: 'Relations entre la Physique Expérimentale et la Physique Mathématique', *Congrès International de Physique de 1900*, **1**, pp. 1–29.

Poincaré, H. [1902]: *La Science et l'Hypothèse.*

Poincaré, H. [1904]: 'L'état actuel et l'avenir de la physique mathematique', *Bulletin des Sciences Mathématics*, **28**, pp. 302–24.

Poincaré, H. [1905]: 'Sur la Dynamique de l'Electron', *Comptes Rendus de l' Academie des Sciences*, **140**, pp. 1504–8.

Poincaré, H. [1906]: 'Sur la Dynamique de l'Electron', *Rendiconti del Circolo Mathematico di Palermo*, **21**, pp. 129–76.

Polanyi, M. [1958]: *Personal Knowledge.*

Popper, K. R. [1959]: *Logic of Scientific Discovery.*

Popper, K. R. [1962]: *Conjectures and Refutations.*

Rayleigh, Lord. [1902]: 'Does motion through the Aether cause Double Refraction?' *Philosophical Magazine*, **4**, pp. 678–83.

REICHENBACH, H. [1924]: *Axiomatik der relativistischen Raum-Zeit-Lehre*. Page references are to the English translation, *Axiomatization of the Theory of Relativity*.

REICHENBACH, H. [1927]: *Philosophie der Raum-Zeit-Lehre*.

RITZ, W. [1908]: 'Récherches critiques sur l'électrodynamique générale', *Ann. Chim. Phys.*, **13**, pp. 145–275.

SALMON, W. C. [1967]: *The Foundations of Scientific Inference*.

SCHAFFNER, K. F. [1968]: "Review of W. C. Salmon's 'The Foundations of Scientific Inference' ", *Isis*, **59**, pp. 447–8.

SCHAFFNER, K. F. [1969]: 'The Lorentz Electron Theory of Relativity', *American Journal of Physics*, **37**, pp. 498–513.

SCHAFFNER, K. F. [1970]: 'Outlines of a Logic of Comparative Theory Evaluation with Special Attention to Pre- and Post-Relativistic Electrodynamics', in R. Stuewer (*ed.*) *Minnesota Studies in The Philosophy of Science*, **5**, pp. 311–54 and 365–73.

SCHAFFNER, K. F. [1972], *Nineteenth-century Aether Theories*.

SCHAFFNER, K. F. [1973]: 'Logic of Discovery and Justification in Regulatory Genetics', *Studies in the History and Philosophy of Science*, **4**, pp. 397–433.

SCHAFFNER, K. F. [1974*a*]: 'Space and Time in Lorentz, Poincaré, and Einstein: Divergent Approaches to the Discovery and Development of the Special Theory of Relativity', in R. Turnbull and P. Machamer (*eds.*): *Space, Matter and Motion*.

SCHAFFNER, K. F. [1974*b*]: 'The Unity of Science and Theory Construction in Molecular Biology', in R. Cohen and R. J. Seeger (*eds.*): *Boston Studies in the Philosophy of Science*, **11**, pp. 207–43.

SEELIG, C. [1956]: *Albert Einstein: A Documentary Biography*.

SHANKLAND, R. S. [1963]: 'Conversations with Albert Einstein', *American Journal of Physics*, **31**, pp. 47–57.

SHANKLAND, R. S. [1964]: 'Michelson-Morley Experiment', *American Journal of Physics*, **32**, pp. 16–35.

SHANKLAND, R. S. [1973]: 'Conversations with Albert Einstein, II', *American Journal of Physics*, **41**, pp. 895–901.

TROUTON, F. T. and NOBLE, H. R. [1903]: 'The Mechanical Forces Acting on a Charged Electric Condensor Moving through Space', *Philosophical Transactions of the Royal Society of London*, **A202**, pp. 165–81.

TOLMAN, R. C. [1910]: 'The Second Postulate of Relativity', *Physical Review*, **31**, pp. 26–40.

VON LAUE, M. [1911]: *Das Relativitätsprinzip*.

WATKINS, J. W. N. [1958]: 'Confirmable and Influential Metaphysics', *Mind*, **67**, pp. 344–65.

WERTHEIMER, M. [1959]: *Productive Thinking*.

WHITTAKER, E. T. [1953]: *A History of the Theories of Aether and Electricity, II*.

WINNIE, J. [1970]: 'Special Relativity without one-way Velocity Assumptions', *Philosophy of Science*, **37**, pp. 81–99 and 223–38.

ZAHAR, E. G. [1973]: 'Why did Einstein's Programme supersede Lorentz's?' *The British Journal for the Philosophy of Science*, **24**, pp. 95–123 and 223–62.

THE TRANSITIVITY OF NON-STANDARD SYNCHRONISMS

1 *Grünbaum's Proof.*
2 *An Exceptional Case.*
3 *Alternative Synchronization Procedures.*
4 *Conclusions.*

In a recent article[1] Adolf Grünbaum has argued that all non-standard synchronisms $0 < \epsilon < 1$ are intransitive. I shall in this note contend that

---

[1] Grünbaum [1969].

Grünbaum's argument, although mathematically flawless, does not prove quite what he has claimed, and I shall argue for what I take to be a correct view of the matter.

I GRÜNBAUM'S PROOF

Let t_1 be the time on a clock at point A at which a light ray is emitted toward point B, and let t_3 be the time on that clock at which the ray returns to point A after instantaneous reflection at point B. The time number, t_2, which a clock at B is set to read when the light ray reaches B is determined by

$$t_2 = t_1 + \epsilon_{AB}(t_3 - t_1), \text{ where } 0 < \epsilon_{AB} < 1.$$

In this equation the particular value of ϵ_{AB} determines the setting of the clock at B required to qualify it as synchronized with the one at A. The standard synchronism is given by the choice $\epsilon_{AB} = \frac{1}{2}$.

Grünbaum's argument rests on three assumptions. First, we suppose that transitivity of synchronism among clocks at points A, B and C is to mean that $\epsilon_{AB} = \epsilon_{BC} = \epsilon_{AC} = \epsilon$. Second, we require that a generalized rule of committed synchronism (*GRCS*) be satisfied; *GRCS* is the principle that at most one act of setting is permissible per clock to achieve the mutual synchronism of any two or more clocks. Finally, we make the empirical assumption, called the two-way light principle, that the round-trip velocity of light measured at a point has the numerical value c *in vacuo*.

The argument then proceeds as follows.[1] Consider three clocks stationed at points A, B and C of an inertial system. Suppose two light rays are emitted from A at $t = 0$ as measured by A's clock. One ray traverses the straight line segment AC of length n; the other traverses the segment AB of length l and then the segment BC of length m, arriving at C later than the first ray. We may assume that the time increment on the clock at C between the arrival of the two rays is

$$\Delta t = \frac{l}{c} + \frac{m}{c} - \frac{n}{c}.$$

Let T_l, T_m and T_n denote the round-trip times of light rays ABA, BCB and ACA as measured on the clocks at A, B and C respectively. By the two-way light principle we know that

$$T_l = \frac{2l}{c},\ T_m = \frac{2m}{c},\ T_n = \frac{2n}{c}.$$

The light ray traveling the path AC reaches C at the time

$$\tau_1 = \epsilon_{AC} T_n = \epsilon_{AC} \frac{2n}{c}.$$

The light ray traveling via B reaches C at the time

$$\tau_2 = \epsilon_{AB} T_l + \epsilon_{BC} T_m = \epsilon_{AB} \frac{2l}{c} + \epsilon_{BC} \frac{2m}{c}.$$

But, since

$$\Delta t = \tau_2 - \tau_1,$$

---

[1] *Ibid.*, pp. 9–11.

by substitution and algebraic manipulation we obtain

$$(1) \qquad \frac{l+m-n}{2} = l\epsilon_{AB}+m\epsilon_{BC}-n\epsilon_{AC}.$$

Imposing the transitivity requirement $\epsilon_{AB} = \epsilon_{BC} = \epsilon_{AC} = \epsilon$ on (*1*) yields

$$\tfrac{1}{2}(l+m-n) = \epsilon(l+m-n),$$

or, since $l+m-n \neq 0$,

$$\epsilon = \tfrac{1}{2},$$

which shows that transitivity of synchronism entails the standard synchronism. On the other hand, if we assume non-standard synchronism, *i.e.* that $\epsilon_{AB} = \epsilon_{BC} = \epsilon_k \neq \frac{1}{2}$, then (*1*) becomes

$$(2) \qquad (l+m)\,(\tfrac{1}{2}-\epsilon_k) = n(\tfrac{1}{2}-\epsilon_{AC}).$$

Equation (*2*) can only hold if either

$$\tfrac{1}{2}-\epsilon_k = \tfrac{1}{2}-\epsilon_{AC} = 0,$$

which is forbidden by the assumption of non-standard synchronism, or

$$\tfrac{1}{2}-\epsilon_k \neq \tfrac{1}{2}-\epsilon_{AC}.$$

Hence, $\epsilon_k \neq \epsilon_{AC}$ and the non-standard synchronism $\epsilon_k \neq \frac{1}{2}$ is intransitive.

2 AN EXCEPTIONAL CASE

What reason could one have for supposing that Grünbaum's proof is not sufficient to establish the intransitivity of all non-standard synchronisms? The reason becomes evident if we reflect on a procedure for synchronizing clocks at collinear points.

Consider three clocks stationed at collinear points A, B and C of an inertial system, where B is between A and C. Suppose a light ray is emitted from A at $t = 0$ as measured by A's clock and travels the distance l from A to B, arriving at B at t_1. At B the ray is partially reflected, the reflected part returning to A at t_2. The transmitted part of the ray travels a distance m from B to C where it is instantaneously reflected at t_3, passes through B at t_4, and arrives at A at t_5. By the two-way light principle we have

$$T_{AB} = \frac{2l}{c},\ T_{BC} = \frac{2m}{c},\ T_{AC} = \frac{2(l+m)}{c}.$$

The synchronization equations may be written as

$$(3) \qquad t_1 = \epsilon_{AB}T_{AB} = \epsilon_{AB}\frac{2l}{c}$$

$$(4) \qquad t_3 = t_1+\epsilon_{BC}T_{BC} = t_1+\epsilon_{BC}\frac{2m}{c}$$

$$(5) \qquad t_3 = \epsilon_{AC}T_{AC} = \epsilon_{AC}\frac{2(l+m)}{c}.$$

Equating (*4*) and (*5*) we obtain

$$(6) \qquad t_1+\epsilon_{BC}\frac{2m}{c} = \epsilon_{AC}\frac{2(l+m)}{c}.$$

Substitution of (*3*) for t_1 in (*6*) and cancellation yields

$$\epsilon_{AB}l+\epsilon_{BC}m = \epsilon_{AC}(l+m),$$

If we now assume that $\epsilon_{AB} = \epsilon_{BC} = \epsilon_k \neq \frac{1}{2}$, we obtain

$$\epsilon_k(l+m) = \epsilon_{AC}(l+m),$$

whence, since $l+m \neq 0$,

$$\epsilon_{AC} = \epsilon_k.$$

This demonstrates the transitivity of all non-standard synchronisms for clocks stationed at collinear points and synchronized as described.

3 ALTERNATIVE SYNCHRONIZATION PROCEDURES

Are we then to conclude that three collinear points constitute a special case not covered by Grünbaum's argument? That this would *not* be the appropriate conclusion is made clear by another simple calculation.

Consider three clocks stationed at collinear points of an inertial system A, B and C, where A is between B and C. Suppose two light rays are emitted from A at time $t = 0$ as measured by A's clock. One ray travels the distance n to C where it arrives at time τ_1; the other ray travels the distance l from A to B where it is instantaneously reflected, and travels then to C where it arrives at time τ_2. It is obvious that this situation represents merely the special case of the proof in *Section* 1 in which $m = l+n$. Substituting $l+n$ for m in (*1*) we get

$$l = \epsilon_{AB}l+\epsilon_{BC}(l+n)-\epsilon_{AC}n. \tag{7}$$

If we now assume that $\epsilon_{AB} = \epsilon_{BC} = \epsilon_k \neq \frac{1}{2}$, equation (*7*) reduces to

$$l+\epsilon_{AC}n = \epsilon_k(2l+n),$$

or

$$n(\epsilon_{AC}-\epsilon_k) = l(2\epsilon_k-1). \tag{8}$$

Since neither l nor n is equal to zero, equation (*8*) can only hold if either

$$\epsilon_{AC}-\epsilon_k = 2\epsilon_k-1 = 0,$$

which is forbidden by the hypothesis of non-standard synchronism, or

$$\epsilon_{AC}-\epsilon_k \neq 0.$$

Hence, $\epsilon_{AC} \neq \epsilon_k$ and the non-standard synchronism $\epsilon_k \neq \frac{1}{2}$ is intransitive. This shows that Grünbaum's argument goes through for the special case of three collinear points.

We may conclude that for three collinear points there are two procedures for clock synchronization. If the two-ray method, the procedure implicit in Grünbaum's argument, is employed, then all non-standard synchronisms are constrained to be intransitive. If the one-ray procedure, the method used in the argument of *Section* 2, is used, then all non-standard synchronisms are required to be transitive.

Moreover, the restriction to collinear points can be removed. We have already seen that Grünbaum's result holds for clocks at any three points. Furthermore, the procedure used in *Section* 2 can also be employed to synchronize clocks at non-collinear points. Consider three non-collinear points A, B and C which form a triangle whose sides AB, BC and AC have lengths l, m and n respectively. Suppose a light ray is emitted from A at $t = 0$ as measured by A's clock and travels to B where it is partially reflected. The reflected part returns directly to A; the transmitted part travels to C and is instantaneously reflected there, then returns to B and thence to A. The argument of *Section 2* applies to this case without alteration, and the resulting non-standard synchronisms are thus required to be transitive. Indeed, the assumption in *Section 2* of collinearity was made only for intuitive concreteness and played no mathematical role in the derivation there, as will have been obvious to the attentive reader.

4 CONCLUSIONS

There are two procedures for synchronizing clocks at three points in an inertial system, and these two procedures impose different constraints on the transitivity properties of non-standard synchronisms. Hence, our conclusions about the transitivity properties of these synchronisms need to be relativized to the synchronization procedure(s) for which they hold. For instance, what Grünbaum has proved is that if the two-ray procedure is used to synchronize clocks at three points then all the non-standard synchronisms will be intransitive. This result is, of course, entirely compatible with my conclusion that if the one-ray procedure is employed to synchronize clocks at three points then all the non-standard synchronisms will be transitive. Whether there are yet other synchronization procedures and what the transitivity properties of non-standard synchronisms are for such procedures remain open questions.[1]

PHILIP L. QUINN
Brown University

REFERENCE

Grünbaum, A. [1969]: 'Simultaneity by Slow Clock Transport in the Special Theory of Relativity', *Philosophy of Science*, 36, pp. 5–43.

[1] I should like to thank Adolf Grünbaum for helpful discussions.

Reviews

BAR-HILLEL, Y. (*ed.*) [1971]: *Pragmatics of Natural Languages.* Dordrecht, Holland: D. Reidel. 17.00$. Pp. vii+231.

This book contains eleven essays which participants at an International Working Symposium held in Jerusalem in 1970 wrote after the Symposium and 'under its impact'. Robert L. Martin makes out a reasonable case for a formal approach to the philosophy of language as one valid method among others, replying to the systematic objections of Katz and Strawson, and arguing that although earlier formal treatments abstracted from pragmatics—that is, from the context of language *use* and from relations between linguistic items, speakers, hearers, writers, and readers—model theory in particular makes possible a formal approach to pragmatics. This must be conceded; but other essays in this volume do little towards realising this possibility. An essay by Leo Apostel seems to develop a semi-formal system and some new jargon for their own sakes; it is not clear that anything else is better understood as a result. The same is true of David Harrah's 'Formal Message Theory' and Helmut Schnelle's 'Language Communication with Children—toward a Theory of Language Use'. Hans-Heinrich Lieb's 'On Subdividing Semiotic' and Asa Kasher's 'A Step toward a Theory of Linguistic Performance' illustrate the dictum that when you have nothing substantive to say you can always propose a new classification.

Niels Christensen's 'Remarks on Logic as Universal Semantics' contains a not very profound attempt to show how 'by philosophical, transcendental reasoning we can argue in favour of at least some universal semantic principles', especially 'the unique position of two-valued logic'. It also gives some fairly obvious arguments against a truth-functional interpretation of the ordinary language connectives 'or', 'and', and 'if-then', but takes no account of possible replies or subtler explanations. L. Jonathan Cohen's 'Some Remarks on Grice's Views about the Logical Particles of Natural Language' does, however, take such account, and this is a well-argued and useful contribution to this specific topic. Cohen contrasts Grice's Conversationalist Hypothesis, that the divergences of the ordinary language connectives from the truth-functional ones are to be explained as resulting from presumptions of normal conversation, with a Semantical Hypothesis, that the differences belong to the basic meaning of these connectives; he argues that while many of the phenomena can be accounted for by either hypothesis, there are some which the Semantical Hypothesis explains better. Two queries, however. First, Cohen is surely wrong to say that Grice's hypothesis 'breathes the same spirit' as speech-act accounts of truth, goodness, and so on: Grice's point is that one can and should ***distinguish*** the linguistic meaning of certain sentences from features that result from their conversational use. Secondly, there is no need to take such constructions, in English dialects and in many other languages, as 'You won't get no beer here' as counter examples to the rule of double negation, as showing that 'not (not *p*)' is not always equivalent to '*p*'. What they contain is not one negation operating upon another, but just one negation multiply represented. (Perhaps for emphasis: compare 'a big big carpet' and 'never never'; or perhaps compare agreements in number and gender between adjectives and nouns, which multiply represent the same semantic element.)

I. M. Schlesinger's 'On Linguistic Competence' also has some genuine content, arguing that judgements about the acceptability of sentences cannot be explained by a theory of grammar alone, but only by a grammar in interaction with psychological principles.

Chaim Perelman's very short essay, 'The New Rhetoric', is mainly a reference to his own extensive work published elsewhere under this title, about the requirements of dialectical argument addressed to an audience (perhaps oneself) as contrasted with those of formal logic. Such an approach is both sound and important; but unfortunately in illustrating the need for it Perelman commits a howler in the very subject he is concerned to develop. He says that 'By ignoring the theory of argumentation, the logicians get a confused view of the *petitio principii*'; to prove the confusion, he quotes remarks from two articles in *The Encyclopedia of Philosophy* and asks 'how can one say both that an argument is formally valid and that it only seems to be valid?' But he should have noted that the articles are by two different authors, so that there is no 'one' who is saying both these incompatible things, and also that one of the two (as it happens, the present reviewer), far from ignoring the theory of argumentation, was expounding, under the heading 'Fallacies in Discourse', an approach equivalent to Perelman's own.

Nathan Stemme's 'Some Aspects of Language Acquisition' develops an empiricist account of how young children could come to understand the grammar of (parts of) a natural language, arguing that the difficulties alleged by 'nativists' like Chomsky disappear if one takes account of semantic factors, especially the 'pairing' or association of recognisably similar linguistic items with recognisably similar situations, features, or reactions. His argument is very much in the spirit of Locke and Hume, and slides, as they do, over problems about *what* we have to pair or associate with various words. I think that Stemme's case is basically sound, but his treatment is not sophisticated enough to shake opposing prejudices. He leaves too many openings for counter-arguments which will be at least rhetorically effective.

As a whole, therefore, this volume is disappointing. Only the essays of Martin, Cohen, Schlesinger, and Stemme have any independent value. The others at most point to work that has been, or might be, done elsewhere. The pragmatics of natural languages is a real subject, but progress can be made in it, as in others, only by observation and experiment guided by ideas or questions, by hard thinking, reasoning, and calculation, or by some combination of these. The 1970 Symposium seems not to have inspired many of its participants to engage in these activities, and without them, as Hume remarked long ago, we shall get nothing but sophistry and illusion.

J. L. MACKIE
University College, Oxford

WINOGRAD, T. [1972]: *Understanding Natural Language*. Edinburgh: Edinburgh University Press. £4.00. Pp. viii+195.

Until very recently, programmes for understanding language have not offered even a pale copy of the understanding of language. Despite their ability to fool their human interlocutors with a frequency which would have delighted the heart of Turing, programmes such as Weizenbaum's ELIZA, Lindsay's SAD SAM, Bobrow's STUDENT, and Raphael's SIR all suffer from a radical flaw in respect of their analogy with human language-users enjoying true comprehension. These programmes, in essence, respond to language rather than understanding it. Their general linguistic strategy is to react in fairly inflexible ways to certain key words, phrases, or syntactic structures. Contextual cues to understanding are crude when they are not entirely lacking. Instead of making intelligent interpretative use of a background system of knowledge ranging over widely varied semantic domains, these programmes employ relatively isolated and inflexible rules to determine their verbal response to linguistic input. In short, such 'speaking' machines do not behave like a man conversing in his native language; rather, they resemble a person resorting to linguistic trickery and semantic sleight-of-hand in order to hide his lack of understanding of a foreign tongue.

Now, with the work of Winograd, we at last have the beginnings of an automatic system which can bring its 'intelligence', or 'reason', to bear in 'interpreting' the language it 'encounters'. (For obvious reasons of aesthetics, scare-quotes will be omitted in the sequel.) The system's resultant understanding of natural language enables it to answer questions, execute commands, and accept information in an interactive English dialogue with a human being. The essential feature of Winograd's programme is that it includes a system of knowledge concerning the universe of discourse, on which the programme can draw in a reasoned fashion so as to make sense of the remarks to which it responds. A flexible interplay between syntactic and semantic features of the linguistic input, and between these and the system's knowledge of the current state and general properties of the world, allows for subtleties of comprehension which vastly outrun the capabilities of earlier assays in the field. Winograd's programme represents a robot which evinces its understanding partly by verbal responses to its interlocutor, and partly by carrying out actions at his command. Access to an internal representation of its own acts in the world—as well as to a model of the world considered independently—enables it to infer the meaning of questions and commands which would otherwise remain obscure.

The potential universe of discourse (and range of possible actions) is stringently limited compared to that of normal conversation. From the point of view of Winograd's programme, the world consists only of the robot itself; a table-top bearing a number of coloured objects and a box; and an interlocutor who asks the robot to move and stack the various objects on the table or in/out of the box, and who plagues it with questions about the world and about its reasons for acting in the way it does. Moreover, there are in fact no objects, no box, no table—and no robot. The table-top world and the computer's actions in it are simulated by internal data processes parallelled by images on a cathode-ray screen. The programme's vocabulary and general knowledge are correspondingly limited.

Within these limitations, however, the power of comprehension evinced by the programme is reasonably impressive, as may be seen from the following fragments of dialogue (the man's remarks are in italics, the machine's in upper case, and Winograd's comments in lower case type):

1. pick up a big red block
OK.

The system answers 'OK' when it carries out a command. In order to pick up the red block, it had to clear it off by finding a space for the green one (initially on top of it) and moving the green one away.

2. grasp the pyramid
I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.
There are three pyramids in the scene, so the programme does nothing.

3. find a block which is taller than the one you are holding and put it into the box.
The system has to figure out what is meant by 'one' and 'it', as well as using deduction to decide which block it is holding.

BY 'IT', I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.

Printed out to keep us informed.

OK.

The hand has to set down the red block to pick up the taller blue one.

4. what does the box contain?
THE BLUE PYRAMID AND THE BLUE BLOCK.

5. what is the pyramid supported by?
Earlier the system responded to 'the pyramid' by saying 'I don't understand which pyramid you mean'. Now, it assumes that we mean the one just mentioned.
THE BOX.

6. how many blocks are not in the box?
Negatives such as 'not' and 'no' are handled whether they appear in clauses or noun groups (such as 'no boxes').
FOUR OF THEM

7. is at least one of them narrower than the one which I told you to pick up?
The phrase 'at least one of them' must be analysed to refer back to the previous question, and 'the one which . . .' must also refer to blocks. Noun groups can have complicated numbers such as 'at least one', 'more than three', 'exactly seven', *etc.* The memory of past events is used to deduce which block is 'the one which I told you to pick up'. Notice the complicated syntactic structure in which 'the one' is actually the object of a clause embedded two levels deeper.
YES, THE RED CUBE.

. . .

17. will you please stack up both of the red blocks and either a green cube or a pyramid?
OK.

. . .

23. had you touched any pyramid before you put the green one on the little cube?
YES, THE GREEN ONE.

The system deduces from its memory what event is being referred to, and can relate other events to it in time. The analysis includes complex tenses of verb groups, and there are special facilities for keeping track of the various forms of irregular verbs like 'have'.

24. *when did you pick it up?*
WHILE I WAS STACKING UP THE RED CUBE, A LARGE RED BLOCK AND A LARGE GREEN CUBE.
A time is described by naming the main goal which was being worked on.

25. *why?*
The system remembers immediately previous sentences to understand questions like this and the following ones.
TO GET RID OF IT.
By keeping track of selected parts of the original subgoal tree, the system has some understanding of its own motives.

26. *why did you do that?*
TO CLEAR OFF THE RED CUBE.
'That' refers to the action mentioned in the answer.

27. *why did you clear off that cube?*
Here, 'that' means the one just referred to.
TO PUT IT ON A LARGE GREEN CUBE.

28. *why did you do that?*
TO STACK UP THE RED CUBE, A LARGE RED BLOCK AND A LARGE GREEN CUBE.

29. *why did you do that?*
BECAUSE YOU ASKED ME TO.
It remembers what it did, not how the request was worded.

. . .

The powerfulness of Winograd's programme, as compared with earlier attempts to simulate the use of natural language, relies upon two basic organisational features. The first is the 'heterarchical' interplay of the various components of the overall programme, while the second is the representation of knowledge as 'procedures', or mini-programmes, rather than as passive items or theorems stored in the data base.

In a heterarchical organisation, the responsibility for control is distributed throughout the system rather than being concentrated at one point. Programmes that are related heterarchically can each call upon the others at many points in their (potentially independent) functioning. The result is an enormous increase in the flexibility of performance (particularly in the sensitivity of control to context) as compared with serial or hierarchical arrangements. Winograd's programme is, in effect, three programmes, concerned respectively with grammar, semantics, and deduction. The first is a parsing system that embodies M. A. K. Halliday's system of systemic grammar, and uses it to recognise the syntactic structure of sentences. The second is a set of semantic programmes dealing with meaning (whether of words, word groups, or whole sentences).

This system is built around a collection of 'semantic specialists' designed to interpret particular syntactic structures such as noun group, adjective group, and so on. And the third is a deductive programme that can solve problems of various kinds and that includes a body of knowledge about the specific 'table-top' universe of discourse. The deductive system is used by the programme at all stages of the analysis, both to direct the parsing process and to deduce facts about the world. Moreover, each of the semantic specialists can use the deductive system and can even call the grammar to do a special bit of parsing before going on with the semantic analysis. In short, the various systems cooperate in subtle ways so as concurrently to interpret the sentences presented to the machine.

The procedural representation of knowledge in Winograd's programme rests on his choice of a version of the recently developed programming language PLANNER for expressing the theorems used by the system. The crucial point about PLANNER theorems is that they are themselves (hierarchically structured) mini-programmes which specify—and, when activated, control—the procedures available for solving problems. A further important aspect of PLANNER is that it is a goal-oriented programming language in which goals can be specified at a high level of generality, without the details of necessary action being specified. In other words, whereas in earlier languages (such as LISP) one has to know precisely what one is looking for in order to ask for it, in PLANNER one can merely state a general *pattern* to which the desired theorem must conform. The difference in problem-solving power and overall flexibility is analogous to the difference between having to tell a telephone mechanic precisely which components he should fix and precisely how he should fix them, and being able to say to him no more than 'Please get me the dialling tone back'. PLANNER therefore enables the programmer to state a problem (and to write a programme which will explore hypothetical plans for its solution) in general terms, without getting bogged down in tactical details. A further significant feature is that theorems expressed in PLANNER can integrally include recommendations as to when and how they should or should not be used. This facility enables the semantic specialists, for instance, to select procedures which are appropriate to the particular case in hand when analysing a potentially ambiguous expression. These selective recommendations expressed within PLANNER theorems allow them (heterarchically) to call each other as and when they are required, rather than in a fixed succession of calls on subordinate/superordinate levels as in the more familiar hierarchical programmes.

It is quite impossible to do justice to the many exciting features of Winograd's programme in a short review, or to give more than a slight indication of the linguistic richness of the interpretative process that it embodies. Even so, Winograd's book is essential reading for anyone interested in the phenomenon of language, whether he be philosopher, psychologist, or linguist, and whether his prime concern be focused on natural or on artificial cognitive systems.

MARGARET A. BODEN
University of Sussex

WHITROW, MAGDA [1971]: *ISIS Cumulative Bibliography. A Bibliography of the History of Science formed from ISIS Critical Bibliographies 1–90, 1913–1965.* London: Mansell Publishing Ltd. in conjunction with the History of Science Society. Pp. lxviii+664 and 789. Volume 1: Personalities A-J; Volume 2: Part 1: Personalities K-Z, Part 2: Institutions.

One had best begin by stating just what these two singularly impressive volumes are and what they are not. Although the present publication grew out of an attempt to satisfy the need for a fifty-year index to *ISIS*, it is not, as is made quite clear in the Foreword, an *index* to the Critical Bibliographies that appeared in *ISIS* during the years in question. Instead, we have to do with the incredibly more useful direct presentation (or more accurately Parts I and II thereof) of the information contained in these Bibliographies as a single, partly reclassified, whole. As the full title of these two volumes indicates, the first two parts of this enterprise reproduced the material in these ISIS Bibliographies bearing upon individuals and institutions, the former constituting by far the larger share of what we have before us.

There are, we are told, some 40,000 items classified under individuals. Where the number of entries under a given individual is large enough, they are subdivided according to 'author entries' (editions and translations of the individual's works) and 'secondary materials' about the individual or about his or her work. In instances in which the entries under a single individual are extraordinarily plentiful, the sub-divisions are multiplied and classify material into, *inter alia*, bibliographical, biographical, forerunners, followers, *etc.* It is also of note that, when the original Critical Bibliography entry contained reference to a review of a book or monograph, such information is duly reported here. Similarly, if the original entry carried annotation, we are here presented at times with a fragment of that annotation, but most frequently with a reference signalling its existence in the initial Critical Bibliography item.

To be sure, no attempt has been made to analyse all Critical Bibliography entries in order to determine whether the item in question contains anything of substance pertinent to a given individual or institution. The classification derives, rather, from (in addition, naturally, to the individual as the author of a text) the occurrence of the name of the individual in question in the title of a book, article, pamphlet, *etc.* (Whenever, incidentally, a number of names occur in a given title, the entry appropriately appears under each individual so named.) Additional entries for an individual or institution have been culled from the annotations in the original Critical Bibliography entry when such annotation has specified such relevance, though the title in question may be silent about it.

For an undertaking of such an enormous scope as the present one, the number of errors seems to be amazingly small. (I have noted but a few mistakes in dating: Aristoxenus should be the fourth century B.C.; Adam of Bocfeld, the thirteenth rather than the fourteenth century, and Johannes de Tinemue (whatever his proper identification might be) also of the thirteenth century; there is also an error that has already received some currency among scholars, entertaining enough in its own way to bear repetition: A. R. Hall's article (volume 2, p. 723) on Robert Merton and science and society in seventeenth century England ('Merton revisited') has been classified under 'Merton College'. (This kind of transformation of Merton into an institution is presumably a bit premature!)

In approaching the *Cumulative Bibliography*, the reader should realise, as the editor emphasises, that the bibliography is in no sense intended as a systematic and complete one. It is selective, the selectivity being very much a function of the particular interests of its founder and primary author: George Sarton. (For Sarton's own views about the virtues of this selectivity and about the Critical Bibliographies as the most valuable part of *ISIS*, one should consult his article in *ISIS*, 41, 1950, pp. 291–8.) Note should be made of the fact, however, that the selective nature of the Critical Bibliographies and, hence, of the present *Cumulative Bibliography*, in no way derives from an overly restricted view of what is important within the history of science. Quite the contrary. Sarton's view of things related the history of science to the history of civilisation or, better, to the history of learning in general, so any number of items of interest are included that bear upon the histories of philosophy, of theology, of society, and upon the philosophy of science. Indeed, so comprehensive was Sarton's view (reflected to a greater extent in the Critical Bibliographies through the early 1950s) that the present editor has found it prudent to omit some material 'bearing on the history of religion, law, economics, general literature, *etc.*'.

Yet just how great a disadvantage is the selective character of the bibliography before us? As a kind of (admittedly inadequate, but perhaps informative) test, I have compared the entries on mathematics and science in Gregor Sebba's systematic *Bibliographia Cartesiana* (The Hague, 1960), for the relevant years with what we have in the *Cumulative Bibliography*. If one sets aside those entries in Sebba that have to do with sections in more general articles or books (histories of physics or of mathematics, for example), the number of items *not* in the *Cumulative Bibliography* is surprisingly low. Indeed, it appears that as many, even more, items are present in the *Cumulative Bibliography* that are *not* in Sebba (apparently Sebba's systematic search for material on Descartes did not extend to a systematic examination of the *ISIS* Critical Bibliographies). On the other hand, if one persists in being hypercritical, there is one area in which the *Cumulative Bibliography* is somewhat weak, an area, one might gather, of considerable interest to readers of the present journal: The history of logic. Thus, of the rather substantial and most significant work done, largely under the impact of the importance of logic within contemporary philosophy and thought, on medieval sources, not only is the greatest part of the secondary material absent, but the editions of the relevant texts of the likes of Abelard, Ockham, Walter Burley, Jean Buridan and William of Sherwood (to mention the most important figures) are also missing. The modern period does not fare much better: There is precious little of the best material on the logic, for example, of Frege, Peano, Schröder, Skolem, *etc.* Furthermore, even when the individual in consideration holds an importance for the history of science outside the area of logic, articles and texts dealing with the latter appear to be systematically under-represented: So it is, for example, in the cases of Aristotle, Leibniz and Bolzano. In view of the fact that Sarton insisted on considering the history of science in its broader aspects, this inattentiveness to the history of logic might be unexpected. Still, whatever reason Sarton may have had for paying less heed to this particular corner of the history of learning, it does represent a weakness in his comprehensive view and, as such, comes through in the *Cumulative Bibliography*.

On balance, however, what is contained in these two volumes in every way offsets any areas or items of minor neglect that they may be charged with. That is to say, although the selective character of the bibliography they contain means that they cannot be utilised as absolutely sufficient and exhaustive in themselves, in every instance material can be found here to supplement almost all other particular bibliographies, no matter how systematic and complete these other bibliographies may purport to be. And this goes for the history of philosophy or intellectual history in general, as well as for the history of science in particular.

To be specific, for the early history of science before (say) 1700, one might well find more entries in *L'Année philologique* (for Antiquity) and in the *Répertoire Bibliographique de la Philosophie*, but these publications have no cumulative representation of the information they contain (although the *Répertoire* does for the years 1934–45 in G. A. De Brie, *Bibliographia Philosophica*, Brussels, 1950) and also lack items that the *Cumulative Bibliography* contains. (One would also wish, incidentally, to supplement pre-1400 material with the bibliography contained in the three volumes of Sarton's *Introduction to the History of Science*, Baltimore, 1927–48.) Similarly, comprehensive special bibliographies devoted to major figures such as Francis Bacon, Descartes, Leibniz, Newton, *etc.*, also consistently can be expanded by appeal to the *Cumulative Bibliography*. Most significant of all, however, are those segments of the more modern stretches of the history of science in which few of even the major protagonists have special bibliographies devoted to their life and work. In such cases—and they comprise a very substantial part of the whole history of science—the volumes at hand provide not merely a supplementary, but the primary, *cache* of bibliographical information.

In sum, the *Cumulative Bibliography* is an indispensable research tool for the history of science proper and for any discipline or particular topic that bears in any way, no matter how slight, upon this history. Given this importance, it is heartening to see that the volumes of it are not being summarily produced, as is all too often the case with bibliographies of a large compass. In the present case, Magda Whitrow, as editor, has given the whole full measure of the proper care and incisive expertise it deserves. The standard she has set will fortunately also be realised in the subsequent volumes that will deal with periods and with subjects within periods. Their appearance will mark another important event for all historical scholarship.

JOHN E. MURDOCH
Harvard University

Brit. J. Phil. Sci. **25** (1974), 92–97 *Printed in Great Britain*

Recent Publications

BOOKS RECEIVED

Acton, H. B. [1973]: *The Illusion of the Epoch*. London: Routledge & Kegan Paul. £1.50. Pp. vii+278.

Adorno, T. W. [1973]: *The Jargon of Authenticity*. London: Routledge & Kegan Paul. £2.75. Pp. xxii+165.

Adorno, T. W. [1973]: *Negative Dialectics*. London: Routledge & Kegan Paul. £4.95. Pp. xxi+416.

Ajdukiewicz, K. [1973]: *Problems and Theories of Philosophy*. Cambridge: Cambridge University Press. £2.90. Pp. xxxiii+170.

Bauman, Z. [1971]: *Culture as Praxis*. London: Routledge & Kegan Paul. £3.25. Pp. v+198.

Beardsley, E. L. and Beardsley, M. C. [1972]: *Invitation to Philosophical Thinking*. New York: Harcourt Brace Jovanovich. £1.35. Pp. viii+178.

Bershady, H. J. [1973]: *Ideology and Social Knowledge*. Oxford: Basil Blackwell. £3.25. Pp. 178.

Beyer, W. R. [1973]: *Hegel-Jahrbuch 1972*. Berlin: Verlag Anton Hain KG, DM 63. Pp. v+334.

Bowlby, J. [1973]: *Separation: Anxiety and Anger*. London: Hogarth Press Ltd. £4.50. Pp. xv+444.

Boyden, A. [1973]: *Perspectives in Zoology*. Oxford: Pergamon Press. £5.50. Pp. xii+288.

Brentano, F. [1973]: *Psychology from an Empirical Standpoint*. London: Routledge & Kegan Paul. £7.00. Pp. xx+415.

Broce, G. [1973]: *History of Anthropology*. Minnesota: Burgess. $1.00. Pp. 43.

Brown, R. [1973]: *Rules and Laws in Sociology*. London: Routledge & Kegan Paul. £3.00 cloth, £2.25 paper. Pp. ix+181.

Brownstein, D. [1973]: *Aspects of the Problem of Universals*. University of Kansas: Lawrence. Pp. 81.

Bulkley, D. H. [1973]: *An Electromagnetic Model for Dynamic Turnover and the Continuous Self-Replication of Biological Macromolecules*. Washington: Rogue Press. $2.00. Pp. 32.

Bunge, M. (*ed.*) [1973]: *Exact Philosophy: Problems, Tools and Goals*. Dordrecht-Holland: D. Reidel Publishing Co. D.fl. 52.00. Pp. vii+214.

Bunge, M. (*ed.*) [1973]: *The Methodological Unity of Science*. Dordrecht-Holland: D. Reidel Publishing Co. D.fl. 45. Pp. viii+264.

Bunge, M. (*ed.*) [1973]: *Philosophy of Physics*. Dordrecht-Holland: D. Reidel Publishing Co. 57 D.fl. Pp. vii+248.

Bunge, M. (*ed.*) [1973]: *Method, Model and Matter*. Dordrecht-Holland: D. Reidel Publishing Co. 45 D.fl. Pp. vi+196.

Castonguay, C. [1972]: *Meaning and Existence in Mathematics*. *Library of Exact Philosophy, Vol. 9*. Vienna, New York: Springer-Verlag. DM. 53.00 Pp. xii+158.

Cattell, R. B. [1973]: *A New Morality from Science: Beyondism*. Oxford: Pergamon Press. £7.25 cloth, £2.95 paper. Pp. xvii+454.

Chihara, C. S. [1973]: *Ontology and the Vicious Circle Principle*. London: Johns Hopkins University Press. £5.65. Pp. xv+257.

Costable, P. [1973]: *Leibniz and Dynamics*. London: Methuen & Co. £3.00. Pp. 141.

Coulter, J. [1973]: *Approaches to Insanity: A Philosophical and Sociological Study*. London: Martin Robertson & Co. £3.50 hardback, £1.65 paperback. Pp. xi+170.

Dahrendorf, R. [1973]: *Homo Sociologicus*. London: Routledge & Kegan Paul. £1.60 cloth, 80p. paper. Pp. vii+99.

Danto, A. C. [1973]: *Analytical Philosophy of Action*. Cambridge: Cambridge University Press. £4.10. Pp. vii+226.

Davis, J. T. [1973]: *The Scientific Approach*. London: Academic Press Ltd. £3.50. Pp. xi+185.

Debus, A. G. (*ed.*) [1972]: *Science, Medicine and Society in the Renaissance* (2 volumes). London: Heinemann Educational Books Ltd. £12.00. Pp. vol. I: 275, vol. II: 338.

Dixon, K. [1973]: *Sociological Theory*. London: Routledge & Kegan Paul. £2.50 cloth, £1.25 paper. Pp. ix+131.

Dummet, M. [1973]: *Frege—Philosophy of Language*. London: Gerald Duckworth & Co. Ltd. £10.00. Pp. xxv+698.

Dymock, W. [1973]: *Parmacographia Indica: A history of the principal drugs of vegetable origin*. Pakistan: The Institute of Health and Tibbi Research. $20. Pp. vii+642.

Eysenck, H. J. and Wilson, G. D. [1973]: *The Experimental Study of Freudian Theories*. London: Methuen & Co. Ltd. £4.90. Pp. xv+405.

Finocchiaro, M. A. [1973]: *History of Science as Explanation*. Detroit: Wayne State University Press. $15.95. Pp. 286.

Foucault, M. [1973]: *The Birth of the Clinic*. London: Tavistock Publications Ltd. £4.20. Pp. xix+215.

Gellner, E. [1973]: *Cause and Meaning in the Social Sciences*. London: Routledge & Kegan Paul. £3.25. Pp. ix+228.

Ghiselin, M. T. [1969]: *The Triumph of the Darwinian Method*. Berkeley, University of California Press. $3.25. Pp. x+287.

Giere, R. and Westfall, R. (*eds.*) [1973]: *Foundations of Scientific Method: The Nineteenth Century*. London: Indiana University Press. £4.40. Pp. ix+306.

Gillies, D. A. [1973]: *An Objective Theory of Probability*. London: Methuen & Co. Ltd. £3.50. Pp. x+250.

Goldmann, L. [1973]: *The Philosophy of the Enlightenment—The Christian Burgess and the Enlightenment*. London: Routledge & Kegan Paul Ltd. £1.50. Pp. ix+100.

Gonseth, F. [1972]: *Time and Method: An Essay on the Methodology of Research*. Illinois: Charles C. Thomas. $16.50. Pp. xiii+453.

Goodman, D. C. (*ed.*) [1973]: *Science and Religious Belief: A Selection of Primary Sources*. London: John Wright & Sons Ltd. £2.75. Pp. 482.

Graham, L. R. [1973]: *Science and Philosophy in the Soviet Union*. London: Allen Lane. £7.00. Pp. xvi+584.

Grandy, R. E. (*ed.*) [1973]: *Theories and Observation in Science*. New York: Prentice-Hall. $7.95 cloth, $3.95 paper. Pp. v+183.

Gross, P. [1973]: *Reflexion, Spontaneität und Interaktion*. Berlin: Friedrich Frommann Verlag. Pp. 163.

Grossmann, R. [1973]: *Ontological Reduction*. Indiana: University Press. $9.00. Pp. 215.

Grünbaum, A. [1973]: *Philosophical Problems of Space and Time* (second, enlarged edition). Dordrecht-Holland: D. Reidel Publishing Co. D.fl. 125 cloth, D.fl. 48 paper. Pp. xxiii+856.

Gusdorf, G. [1973]: *L'avenement des Sciences Humaines au Siecle des Lumieres* (les sciences humaines et la conscience occidentale—VI). Paris: Payot. Pp. 589.

Hallam, A. [1973]: *A Revolution in the Earth Sciences: From Continental Drift to Plate Tectonics*. Oxford: Clarendon Press. £4.00. Pp. ix+127.

Handy, R. and Harwood, E. C. [1973]: *Useful Procedures of Inquiry*. London: Behavioral Research Council. $12.50. Pp. vii+225.

Harrison, R. [1973]: *Warfare*. New York: Burgess Publishing Co. $1.40. Pp. iii+61.

Hawkes, J. [1973]: *Nothing But or Something More*. London and Seattle: University of Washington Press. £1.75. Pp. xi+30.

Hayek, F. A. [1973]: *Law, Legislation and Liberty, Vol. I Rules and Order*. London: Routledge & Kegan Paul. £3.00. Pp. xi+184.

Hegel, G. W. F. [1973]: *Vorlesungen über Rechtsphilosophie*. Stuttgart: Frommann-Holzboog. DM 198. Pp. 126.

Hegenberg, L. [1973]: *Logica o Calculo de Predicados* (2 volumes). Sao Paulo: Editora Herder. Pp. xi+226.

Herbst, P. G. [1973]: *Behavioural Worlds: The Study of Single Cases*. London: Tavistock Publications Ltd. £1.50. Pp. xiv+248.

Hewes, G. W. [1973]: *The Origin of Man*. Minneapolis, Minnesota: Burgess Publishing Co. $1.25. Pp. iii+60.

Hooker, C. A. (*ed.*) [1973]: *Contemporary Research in the Foundations and Philosophy of*

Quantum Theory, Vol. 2. Dordrecht-Holland: D. Reidel Publishing Co. D.fl. 56 Pp. xx+385.

Hooykaas, R. [1973]: *Religion and the Rise of Modern Science*. Scottish Academic Press. Hardback £2.25, paperback £1.25. Pp. ix+162.

Hull, D. L. [1973]: *Darwin and his Critics*. Harvard: University Press. £9.25. Pp. xii+473.

Huxley, J. [1973]: *Memories II*. London: George Allen & Unwin Ltd. £4.95. Pp. 269.

Jacob, F. [1973]: *The Logic of Living Systems*. London: Allen Lane. £4.00. Pp. viii+348.

Jaki, S. L. [1973]: *The Milky Way: An Elusive Road for Science*. Newton Abbot: David and Charles. £5.95. Pp. xi+252.

Janich, P. [1973]: *Zweck Und Methode der Physik aus Philosophischer Sicht*. Konstanzer Universitätsreden. DM.11.40. Pp. 36.

Janik, A. and Toulmin, S. [1973]: *Wittgenstein's Vienna*. London: Weidenfeld & Nicolson. £5.25. Pp. 314.

Jarvie, I. C. [1973]: *Functionalism*. Minneapolis, Minnesota: Burgess Publishing Co. $1.00. Pp. 39.

Jaspers, K. [1973]: *Philosophie Vol. I, Philosophische Weltorientierung*. Berlin: Springer-Verlag. $7.20. Pp. lv+340.

Jaspers, K. [1973]: *Philosophie Vol. II, Existenzerhellung*. Berlin: Springer-Verlag. $7.20. Pp. xi+440.

Jaspers, K. [1973]: *Philosophie Vol. III, Metaphysik*. Berlin: Springer-Verlag. $7.20. Pp. viii+276.

Jauch, J. M. [1973]: *Are Quanta Real? A Galilean Dialogue*. Indiana: University Press. £3.30. Pp. xii+106.

Jech, T. J. [1973]: *The Axiom of Choice*. Amsterdam: North-Holland Publishing Co. D.fl. 40.00. Pp. vii+202.

Jeffreys, Sir H. [1973]: *Scientific Inference*. Cambridge: Cambridge University Press. £5.50. Pp. viii+273.

Jensen, A. R. [1973]: *Educability and Group Differences*. London: Methuen & Co. £3.90. Pp. xiii+407.

Jensen, A. R. [1973]: *Educational Differences*. London: Methuen & Co. £5.75. Pp. 462.

Kalocsai, D. [1973]: *Le Probleme des Regles de la Morale 'Provisoire' de Descartes*. £2.20. Pp. 110.

Kambartel, F. and Mittelstrass, J. (*eds.*) [1973]: *Zum Normativen Fundament der Wissenschaft*. Frankfurt: Athenäum Verlag. Pp. x+330.

Kaplan, M. A. [1973]: *Strategic Thinking and its Moral Implications*. Chicago: The University of Chicago Center for Policy Study. $9.75 cloth, $2.95 paper. Pp. 155.

Kierkegaard, S. [1973]: *La Neutralita' Armata*. Messina: A. M. Sortino Editore. Lire 3,500. Pp. 210.

Kline, M. [1973]: *Mathematical Thought from Ancient to Modern Times*. London: Oxford University Press. £12.00. Pp. xvii+1238.

Kluver, J. [1973]: *Operationalismus Kritik und Geschichte einer Philosophie der exakten Wissenschaften*. Stuttgart: Friedrich Frommann Verlag. Pp. 220.

Koningsveld, H. [1973]: *Empirical Laws, Regularity and Necessity*. Wageningen, Holland: H. Veenman & B. V. Zonen. Pp. 129.

Koyré, A. [1973]: *The Astronomical Revolution*. London: Methuen & Co. Ltd. £6.50. Pp. 531.

Kruger, L. [1973]: *Der Begriff des Empirismus*. Berlin, New York: Walter de Gruyter & Co. DM. 68.00. Pp. xii+283.

Kutschera, F. V. [1973]: *Wissenschaftstheorie I and II*. Munich: Wilhelm Fink Verlag. DM. 19.80. Pp. 296.

Lackey, D. (*ed.*) [1973]: *Bertrand Russell: Essays in Analysis*. London: George Allen & Unwin Ltd. £5.50. Pp. 345.

Landé, A. [1973]: *Quantum Mechanics in a New Key*. New York: Exposition Press. $6.50 Pp. vii+131.

Laszlo, E. (*ed.*) [1973]: *The World System*. George Braziller, Inc. $7.95 cloth, $2.95 paper. Pp. x+215.

Leblanc, H. (*ed.*) [1973]: *Truth, Syntax and Modality*. Amsterdam: North-Holland Publishing Co. $21.10. Pp. vii+317.

Lewis, D. [1973]: *Counterfactuals.* Harvard: Harvard University Press. $7.50. Pp. x+150.

Lindsay, R. B. [1973]: *Julius Robert Mayer: Prophet of Energy.* Oxford: Pergamon Press Ltd. £2.75. Pp. viii+238.

Lucas, J. R. [1973]: *A Treatise on Time and Space.* London: Methuen; New York: Barnes & Noble. £7.00. Pp. xi+321.

Meehl, P. E. [1973]: *Psychodiagnosis: Selected Papers.* Minneapolis: University of Minnesota Press. £7.25. Pp. xxii+359.

Mehra, J. (*ed.*) [1973]: *The Physicist's Conception of Nature.* Dordrecht-Holland: D. Reidel Publishing Co. D.fl. 225. Pp. xxiii+839.

Melehy, M. A. [1973]: *Foundations of the Thermodynamic Theory of Generalized Fields.* Baltimore: Mono Book Corporation. $15.00. Pp. xxiii+255.

Melhuish, G. [1973]: *The Paradoxical Nature of Reality.* St. Vincent's Press. £2.90. Pp. xix+196.

Mellor, D. H. [1971]: *The Matter of Chance.* Cambridge: Cambridge University Press. £4.00. Pp. xiii+190.

Mendelssohn, K. [1973]: *The World of Walther Nernst: The Rise and Fall of German Science.* London: Macmillan. £4.95. Pp. viii+191.

Michaelis, A. R. and Harvey, H. (*eds.*) [1973]: *Scientists in Search of their Conscience.* Berlin: Springer-Verlag. DM. 46, $16.30. Pp. xiii+230.

Michel, P-H. [1973]: *The Cosmology of Giordano Bruno.* London: Methuen & Co. £4.50. Pp. 306.

Mitcham, C. and Mackey, R. [1972]: *Philosophy and Technology.* Free Press. £5.85. Pp. ix+378.

Nakayama, S. and Sivin, N. (*eds.*) [1973]: *Chinese Science: Explorations of an Ancient Tradition.* Cambridge, Massachusetts: The MIT Press. $12.50. Pp. xxxvi+334.

Neurath, O. [1973]: *Empiricism and Sociology.* Dordrecht-Holland: D. Reidel Publishing Co. D.fl. 52. Pp. xiv+473.

Niiniluoto, I. and Tuomela, R. [1973]: *Theoretical Concepts and Hypothetico-Inductive Inference.* Dordrecht-Holland: D. Reidel Publishing Co. D.fl. 70. Pp. x+264.

Nørreklit, L. [1973]: *Concepts.* Odense: Odense University Press. Dan.Kr.70. Pp. 226.

Olson, R. G. [1969]: *Meaning and Argument.* London: Harcourt Brace Jovanovich Ltd. £4.15. Pp. xi+419.

Orange, A. D. [1973]: *Philosophers and Provincials: The Yorkshire Philosophical Society from 1822 to 1844.* York: Yorkshire Philosophical Society. £1.75. Pp. 76.

Ostrowski, J. J. [1973]: *Alfred Espinas, Precurseur de la Praxeologie.* Paris: R. Pichon et R. Durand-Aurzias, 80 F. Pp. 394.

Pattee, H. H. (*ed.*) [1973]: *Hierarchy Theory: The Challenge of Complex Systems.* New York: George Braziller, Inc. $6.95 cloth, $2.95 paper. Pp. vii+156.

Pendergast, R. J. [1973]: *Cosmos.* New York: Fordham University Press. $9.00. Pp. xiii.

Peters, R. S. (*ed.*) [1973]: *The Philosophy of Education.* Oxford: Oxford University Press. £1.10. Pp. 278.

Poignant, R. [1973]: *Education in the Industrialised Countries.* Netherlands: Martinus Nijhoff. 52.50 D.fl. Pp. xi+324.

Polanyi, M. [1973]: *Personal Knowledge—Towards a Post-Critical Philosophy.* London: Routledge & Kegan Paul Ltd. £2.00. Pp. xiv+428.

Polten, E. P. [1973]: *Critique of the Psycho-physical Identity Theory.* The Hague: Mouton & Co. D.fl. 34. Pp. xiv+290.

Reitdijk, C. W. [1973]: *On the Explanation of Matter Wave Interference: towards the end of indeterministic physics.* Assen, Holland: Royal Vangorcum Ltd. D.fl. 26. Pp. xii+98.

Remmling, G. W. (*ed.*) [1973]: *Towards the Sociology of Knowledge: origin and development of a sociological thought style.* London: Routledge & Kegan Paul. £6.25. Pp. xxi+463.

Rescher, N. [1973]: *The Coherence Theory of Truth.* London: Clarendon Press; Oxford University Press. £4.50. Pp. xiv+374.

Rescher, N. [1974]: *Conceptual Idealism.* Oxford: Blackwell. £3.95. Pp. xiii+204.

Ricoeur, P. [1972]: *L'Ermeneutica del Sublime Saggi per una Critica dell' Illusione.* Messina: A. M. Sortino Editore. Pp. 204.

Riepe, D. (*ed.*) [1973]: *Phenomenology and Natural Existence.* New York: State University of New York Press. $15.00. Pp. viii+408.

Robinet, A. [1973]: *Le Defi Cybernetique*. Paris: Gallimard. Pp. 232+7.

Roche, M. [1973]: *Phenomenology, Language and the Social Sciences*. London: Routledge & Kegan Paul Ltd. £5.00. Pp. vii+361.

Roslansky, J. D. (*ed.*) [1973]: *The End of Life*. Amsterdam: North Holland Publishing Co. $8.80. Pp. 83.

Ruse, M. [1973]: *The Philosophy of Biology*. London: Hutchinson University Library. £3.00 cloth, £1.84 paper. Pp. 231.

Russell, C. A. (*ed.*) [1973]: *Science and Religious Belief, a Selection of Recent Historical Studies*. London: University of London Press Ltd. £3.40 cloth, £1.70 paper. Pp. viii+323.

Ryan, A. (*ed.*) [1973]: *The Philosophy of Social Explanation*. Oxford: Oxford University Press. 90p. Pp. 228.

Sachs, M. [1973]: *The Field Concept in Contemporary Science*. Illinois: Charles C. Thomas. $9.75. Pp. xi+120.

Sagan, C. and Page, T. [1973]: *UFO's: A Scientific Debate*. London: Cornell University Press. £5.65. Pp. xxxi+310.

Salmon, W. C. [1973]: *Logic* (2nd edition). New Jersey: Prentice-Hall Inc. £2.95. Pp. x+150.

Salomon, J-J. [1973]: *Science and Politics*. London: Macmillan. £5.50. Pp. xxii+257.

Schächter, J. [1973]: *Prolegomena to a Critical Grammar*. Dordrecht-Holland: D. Reidel Publishing Co. D.fl. 23. Pp. xx+164.

Scheffler, I. [1973]: *Reason and Teaching*. London: Routledge and Kegan Paul. £3.50. Pp. xi+203.

Scheibe, E. [1973]: *The Logical Analysis of Quantum Mechanics*. Oxford: Pergamon Press. Pp. vii+204.

Scheibe, E. and Sussman, G. (*eds.*) [1973]: *Einheit und Vielheit, Festschrift für Carl Friedrich v. Weizsäcker zum 60 Geburtstag*. Göttingen: Vandenhoech & Ruprecht. DM. 75. Pp. 304.

Seeger, R. J. [1973]: *Benjamin Franklin: New World Physicist*. Oxford: Pergamon Press. £2.50. Pp. xii+190.

Shackle, G. L. S. [1973]: *An Economist Querist*. Cambridge: Cambridge University Press. £2.50. Pp. vii+135.

Shea, W. R. [1972]: *Galileo's Intellectual Evolution*. London: Macmillan Press. £4.95. Pp. vii+204.

Sjobring, H. [1973]: *Personality, Structure and Development*. Copenhagen: Munksgaard. Pp. 204.

Smith, A. D. [1973]: *The Concept of Social Change*. London: Routledge & Kegan Paul. £3.75 cloth, £1.95 paper. Pp. ix+198.

Smith, A. [1973]: *Four Uses of Heuristics*. London: Panparameter. £2.50. Pp. 207.

Solomon, J. [1973]: *The Structure of Matter*. Newton Abbott: David & Charles. £3.95 cloth, £1.95 paper. Pp. 179.

Solomon, J. [1973]: *The Structure of Space*. Newton Abbott: David & Charles. £4.50 cloth, £2.25 paper. Pp. 219.

Specht, R. [1973]: *Innovation und Folgelast*. Stuttgart: Frommann-Holzboog. DM. 22. Pp. 237.

Spier, R. F. G. [1973]: *Material Culture and Technology*. Minnesota: Burgess. $0.75. Pp. iii+36.

Spisani, F. [1972]: *The Meaning and Structure of Time*. Bologna: Azzoguidi. Lire 2,500. Pp. 161.

Stamper, R. [1973]: *Information in Business and Administrative Systems*. London: B. T. Batsford. £5.50. Pp. 362.

Stegmuller, W. [1973]: *Personelle und Statistische Wahrscheinlichkeit* (*IV erster halbband*). Springer-Verlag. $25.20. Pp. xv+420.

Stegmuller, W. [1973]: *Personelle und Statistische Wahrscheinlichkeit* (*IV zweiter halbband*). Springer-Verlag. $31.40. Pp. xxii+558.

Stegmuller, W. [1973]: *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Band 2 Theorie und Erfahrung*. Springer-Verlag. DM. 56; US $23.00.

Stubbe, H. [1973]: *History of Genetics from Prehistoric Times to the Rediscovery of Mendel's Laws*. Massachusetts: MIT Press. $14.95. Pp. x+356.

Sutherland, J. W. [1973]: *A General Systems Philosophy for the Social and Behavioral Sciences.* New York: George Braziller. $6.96 cloth, $2.95 paper. Pp. ix+210.

Swinburne, R. (*ed.*) [1974]: *The Justification of Induction.* Oxford: Oxford University Press. 75p. paperback. Pp. 179.

Taylor, Sir James [1973]: *The Scientific Community.* Oxford: Oxford University Press. £1.05. Pp. viii+79.

Teich, M. and Young, R. [1973]: *Changing Perspectives in the History of Science: Essays in Honour of Joseph Needham.* London: Heinemann. £6.50. Pp. xxi+490.

Tomasino, R. [1973]: *Segno E Materialsimo.* Lire 1,600. Pp. iii+125.

Trigg, R. [1973]: *Reason and Commitment.* Cambridge: Cambridge University Press. £3.20. Pp. 172.

Tuomela, R. [1973]: *Theoretical Concepts.* Berlin: Springer-Verlag. $5.10. Pp. xiv+254.

Von Holst, E. [1973]: *The Behavioural Physiology of Animals and Man. Selected Papers, Volume 1.* London: Methuen & Co. £6.50. Pp. xiv+341.

Vorzimmer, P. J. [1973]: *Charles Darwin: The Years of Controversy.* London: University of London Press Ltd. £4.40. Pp. xix+300.

Waardenburg, J. [1973]: *Classical Approaches to the Study of Religion, Vol. 1: Introduction and anthology.* The Hague: Mouton & Co. D.fl.50. Pp. xiv+742.

Wang, H. [1974]: *From Mathematics to Philosophy.* London: Routledge & Kegan Paul. £7.75.

Watkins, J. W. N. [1973]: *Hobbes's System of Ideas.* London: Hutchinson University Library. £1.75. Pp. vi+132.

Westfall, R. S. [1973]: *Science and Religion in Seventeenth Century England.* Michigan: University of Michigan Press. $2.95. Pp. x+239.

White, L. A. and Dillingham, B. [1973]: *The Concept of Culture.* Minnesota: Burgess Publishing Co. $1.50. Pp. iii+73.

White, M. [1973]: *Pragmatism and the American Mind: Essays and Reviews in Philosophy and Intellectual History.* New York: Oxford University Press. £4.00. Pp. xi+265.

Whitley, R. (*ed.*) [1974]: *Social Processes of Scientific Development.* London: Routledge & Kegan Paul. £5.95. Pp. ix+286.

Williams, R. H. (*ed.*) [1973]: *To Live and To Die: When, Why and How.* Berlin: Springer-Verlag. $12.95. Pp. xix+346.

Winograd, T. [1972]: *Understanding Natural Language.* Edinburgh: Edinburgh University Press. £4.00. Pp. vii+195.

Wrighton, R. F. [1974]: *Elementary Principles of Probability and Information.* London: Academic Press. £1.95. Pp. viii+91.

Zeeman, J. J. [1973]: *Modal Logic.* Oxford: Clarendon Press. £6.50. Pp. ii+289.

Zinov'ev, A. A. [1973]: *Foundations of the Logical Theory of Scientific Knowledge.* Dordrecht-Holland: D. Reidel Publishing Co. D.fl. 43.00 paper, D.fl. 72.00 cloth. Pp. xxii+300.

THE CHARLES LYELL CENTENARY SYMPOSIUM

The sixth symposium of the International Committee on the History of Geological Sciences will be held in the first week of September 1975 at Imperial College, London.

The Symposium will explore Lyell's work in the context of 19th century science, its background and the origin of his ideas, and its legacy for modern science.

There will be excursions to Scotland and Hampshire associated with the Symposium.

Copies of the First Notice can be obtained from the Secretary of the Organising Committee: J. C. Thackray, Institute of Geological Sciences, Exhibition Road, London SW7 2DE.

HISTORY OF SCIENCE

*Is the **only** review of literature and research in the history of science, medicine and technology in its intellectual and social context*

FOUNDED IN 1962 AS AN ANNUAL VOLUME, *History of Science* discusses problems and sources for future research in the history of science, medicine and technology, and surveys recent publications. It now appears each year in quarterly parts of about 88 pages each.

VOLUME 11 (1973) included

Robert G. Frank *Science, Medicine and the Universities of Early Modern England*

J. E. McGuire *Newton and the Demonic Furies*

Charles B. Schmitt *Towards a Reassessment of Renaissance Aristotelianism*

Victor E. Thoren *Tycho Brahe: Past and Future Research*

Roger Smith *The Background of Physiological Psychology in Natural Philosophy*

P. M. Heimann *Newtonian Natural Philosophy and the Scientific Revolution*

together with ten essay reviews and supporting material.

THE ANNUAL SUBSCRIPTION is £8 post free outside the Americas and Japan. Readers who prefer to receive the four issues in a bound volume each December may do so for the same subscription. New subscribers may purchase the first eleven volumes (bound) for £38 post free, *including* the subscription to Volume 12 (1974) *and,* **as a free gift**, the collected papers of Lord Rutherford in three large volumes. Write to: Circulation Manager, *History of Science*, Whipple Museum, Free School Lane, Cambridge CB2 3RH.

29 OCT 1976

Australasian Journal of Philosophy

Editor: Robert Brown

Published in May, August and December

Subscriptions are $8 (Aust.) for institutions and $5 (Aust.) for individuals. They should be sent to the Business Manager at the *Department of Philosophy, School of General Studies, A.N.U., Canberra, A.C.T., 2600.*

Manuscripts should be sent to the Editor at the *Department of Philosophy, Research School of Social Sciences, A.N.U., Canberra, A.C.T., 2600.*

Noûs

Nihil philosophici a nobis alienum putamus

A Quarterly Journal of Philosophy

edited by Hector-Neri Castañeda

NOÛS publishes quality essays and brief discussions on philosophical problems regardless of the author's philosophical school or point of view. From time to time, NOÛS will publish symposia, surveys of work recently done on specially selected philosophical topics, critical studies of recent books (rather than reviews), and catalogues of recent publications in philosophy.

A quarterly, NOÛS is published in March, May, September, and November each year. For individuals the subscription rates are: one year, $8.00; two years, $15.00; three years, $22.00. For *bona fide* students: $6.00 per year. For institutions: $12.00 per year. Add 45 cents per year for foreign subscriptions.

VOLUME VIII NUMBER 1 MARCH 1974

CONTENTS

SYMPOSIA PAPERS to be read at the meeting of the Western Division of the American Philosophical Association in St. Louis, Missouri, April 27–29, 1974.

CHARLES PARSONS—Sets and Classes

DAVID BRAYBROOKE—From Economics to Aesthetics: The Rectification of Preferences

HILARY PUTNAM—The Refutation of Conventionalism

JAEGWON KIM—Noncausal Connections

DAVID GAUTHIER—Rational Cooperation

REINHARDT GROSSMANN—Meinong's Doctrine of the *Aussersein* of the Pure Object

Recent Publications in Philosophy

Order from

The Secretary, NOÛS, Department of Philosophy
702 Ballantine Hall, Indiana University
Bloomington, Indiana 47401

AMERICAN
PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

VOLUME 11/NUMBER 3 JULY 1974

CONTENTS

I WILLIAM A. PARENT: *Recent Work on the Concept of Liberty*
II DON LOCKE: *Reasons, Wants and Causes*
III J. MICHAEL YOUNG: *The Ontological Argument and the Concept of Substance*
IV JOYCE TREBILCOT: *Aprudentalism*
V JOHN FISHER: *Universalizability and Judgments of Taste*
VI RUTH BARCAN MARCUS: *Classes, Collections and Individuals*
VII STEPHEN SCOTT: *Rituals of 'Ought'*
VIII ROBERT YOUNG: *Petitioning God*

Annual Subscription: $18.00 to Institutions (*ex*cluding Annual Monograph, price $7.00)
$8.00 to Individuals (*in*cluding Annual Monograph)

Published by BASIL BLACKWELL
Orders to: 108 Cowley Road, Oxford OX4 1JF, England

Philosophy and Phenomenological Research
AN INTERNATIONAL QUARTERLY
Edited by MARVIN FARBER

MARCH, 1974, VOLUME XXXIV, No. 3

CONTENTS

I. ARTICLES
Personality, by REYNOLD LAWRIE
Phenomenology and the Form of the Novel: Toward an Expanded Critical Method, by DABNEY W. TOWNSEND
Can Sartre Count?, by FRED KERSTEN
Commodities and Exchange: Notes for an Interpretation of Marx, by ANATOLE ANTON
The Frege-Hilbert Controversy, by MICHAEL DAVID RESNIK
Linguistic Analysis and Epistemic Encounters, by E. M. ADAMS
Reasons as Causal Explanations, by R. C. SOLOMON
II. DISCUSSIONS, REVIEWS
III. COMMUNICATIONS, NOTES AND NEWS, RECENT_PUBLICATIONS

Subscription prices $7.50 per year for individuals; $9.00 per year for institutions. Separate issues, $2.00 and $2.50 respectively. Postage and handling for back issues only. Papers submitted for publication will not be returned unless accompanied by a self-addressed stamped envelope, or equivalent postage.

A Circular listing the main contents of the Journal since 1940 will be sent upon request.

Address all communications to
PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH
State University of New York at Buffalo, Buffalo, New York, 14226

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1974-5

President: Professor J. W. N. Watkins
President-Elect: Professor D. J. O'Connor
Hon. Secretary: Dr P. M. Williams
Hon. Treasurer: Dr D. A. Gillies
Members: Dr I. Hacking, Mr H. Martins, Dr M. Machover, Mr N. Maxwell, Mr W. Newton-Smith, Dr E. Zahar

EDITORIAL BOARD OF THE JOURNAL

Professor R. B. Braithwaite
Professor H. Dingle
Professor Sir John Eccles
Professor Sir Harold Jeffreys
Professor W. C. Kneale
Professor S. Körner
Professor M. Polanyi
Professor Sir Karl Popper
Professor G. J. Whitrow
Professor J. H. Woodger

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules:

'The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all persons who have been proposed by two members, approved by the Committee, and elected at an ordinary meeting. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £1·00 ($2.50). Those who wish to join should write to the Hon. Secretary, Dr P. M. Williams, School of Mathematical and Physical Sciences, The University of Sussex, Falmer, Brighton, BN1 9QH.

Brit. J. Phil. Sci. Vol. 25 No. 1 March 1974

The British Journal for the Philosophy of Science

Articles

1–23 ALAN E. MUSGRAVE Logical versus Historical Theories of Confirmation.

Discussions

25–28 PAUL K. FEYERABEND Zahar on Einstein.
29–45 ARTHUR I. MILLER On Lorentz's Methodology.
45–78 KENNETH F. SCHAFFNER Einstein versus Lorentz: Research Programmes and the Logic of Theory Evaluation.
78–82 PHILIP L. QUINN The Transitivity of Non-Standard Synchronisms.

Reviews

83–91 J. L. MACKIE on Y. Bar-Hillel (*ed.*): *Pragmatics of Natural Languages;* MARGARET A. BODEN on T. Winograd: *Understanding Natural Language;* JOHN E. MURDOCH on M. Whitrow: *Isis Cumulative Bibliography.*

Recent Publications

Programmes of Meetings

29 OCT 1976

Printed in Great Britain at the University Press, Aberdeen

Volume 25 Number 2 June 1974
BJPIA5 25 (No. 2) 99-208 (1974) 0007-0882

The British Journal for the Philosophy of Science

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Acting Editor: JOHN WORRALL

Assistant Editors: COLIN HOWSON AND ELIE ZAHAR

Subscriptions
The *Journal* is published quarterly. The subscription price is £7.00 net (U.S.A. and Canada U.S. $22.00) per volume post free; single issues £2.50 net (U.S.A. and Canada U.S. $8.00) plus postage. The subscription is £4.50 (U.S.A. and Canada U.S. $14.00) for members of the British Society for the Philosophy of Science, who should order direct from the publisher. (Membership of the Society is international and is by no means confined to people resident in the United Kingdom.) Orders may be sent to any bookseller or direct to Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT.

Notes to Contributors
Articles, correspondence and books for review should be sent to the Acting Editor, John Worrall at the Department of Philosophy, London School of Economics, Houghton Street, Aldwych, London WC2A 2AE. Articles should be accompanied by abstracts of about 100 words.

Material is accepted for publication in the *Journal* on the condition that copyright is assigned to the British Society for the Philosophy of Science.

Three copies of all contributions should be sent, typed on one side only with wide margins and double-spaced. Footnotes, which should be as few as possible, should be typed separately, double-spaced at the end of the text. They should be referred to by consecutive numbering throughout the text, in which they will be inserted in page proof. Bibliographical references should be listed alphabetically at the end. The style of references should follow current *Journal* practice.

Long quotations are printed in small type without quotes, and should be so marked in the typescript. Single quotes should be used to form the names of words and sentences, and double quotes for all other purposes. Diagrams, tables, and illustrations should be on separate sheets, with their desired position in the text clearly indicated. Symbolic formulae appearing as separate lines should be clearly distinguished in the text, and foreign letters (Greek, German, etc.) indicated in the margin. Material to be printed in italics, especially English letters used as symbols ('*A*', '*x*', etc.), should be underlined. On matters of style and layout, authors should consult CHAUNDY, T. W. *et al.* [1964]: *The Printing of Mathematics*, London: Oxford University Press.

Offprints of articles may be ordered when corrected galley proofs are returned. Fifty are supplied free ; the price of extra copies is given on the order form.

© British Society for the Philosophy of Science 1974

For permission to reproduce material from *The British Journal for the Philosophy of Science*, please apply to the Editor.

Back Volumes
Volumes 1-16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17-24 may be obtained from Cambridge University Press.

The British Journal for the Philosophy of Science

Volume 25

CAMBRIDGE UNIVERSITY PRESS 1974

CONTENTS

ARTICLES, DISCUSSIONS, MISCELLANEOUS

ARTICLES

page

1 LOGICAL VERSUS HISTORICAL THEORIES OF CONFIRMATION
Alan Musgrave

99 PROGRESS AND DEGENERATION IN THE 'I.Q. DEBATE' (I)
Peter Urbach

137 PATTERN RECOGNITION IN COMPUTERS AND THE HUMAN BRAIN
Roland Puccetti

209 THE UNPREDICTABILITY OF FREE CHOICES
George Schlesinger

223 PROPOSITIONS, POSSIBLE LANGUAGES AND THE LIAR'S REVENGE
John Post

235 PROGRESS AND DEGENERATION IN THE 'I.Q. DEBATE' (II)
Peter Urbach

301 UNIFORMITY AND INDUCTION
John C. Graves

DISCUSSIONS

25 ZAHAR ON EINSTEIN
Paul K. Feyerabend

29 ON LORENTZ'S METHODOLOGY
Arthur I. Miller

45 EINSTEIN VERSUS LORENTZ: RESEARCH PROGRAMMES AND THE LOGIC OF THEORY EVALUATION
Kenneth F. Schaffner

78 THE TRANSITIVITY OF NON-STANDARD SYNCHRONISMS
Philip L. Quinn

155 ON POPPER'S DEFINITIONS OF VERISIMILITUDE
Pavel Tichý

160 POPPER'S DEFINITIONS OF VERISIMILITUDE
John H. Harris

page

166 POPPER'S QUALITATIVE THEORY OF VERISIMILITUDE
David Miller

178 ON THE COMPARISON OF FALSE THEORIES BY THEIR BASES
David Miller

261 ON THE INFIRMITIES OF GILLIES'S RULE
Stephen Spielman

265 ON NEYMAN'S PARADOX AND THE THEORY OF STATISTICAL TESTS
M. L. G. Redhead

271 SITUATIONAL DETERMINISM IN ECONOMICS
Fritz Machlup

285 SITUATIONAL DETERMINISM IN ECONOMICS: THE IMPLICATIONS OF LATSIS'S ARGUMENT FOR THE HISTORIAN OF ECONOMICS
A. W. Coats

289 A CORRECTION CONCERNING COMPLEXITY
I. J. Good

319 POPPER AND QUANTUM LOGIC
Erhard Scheibe

329 QUANTUM FACT AND FICTION
D. K. Nartonis

332 ARE THE 'MEMBERS' OF BIOLOGICAL SPECIES 'SIMILAR' TO EACH OTHER?
David L. Hull

334 PROFESSOR HULL AND THE EVOLUTION OF SCIENCE
L. Jonathan Cohen

336 DID EINSTEIN'S PROGRAMME SUPERSEDE LORENTZ'S?
S. J. Prokhovnik

340 A LITTLE LEARNING CAN BE DANGEROUS
I. J. Good

REVIEW ARTICLES

343 OTTO NEURATH (review of O. Neurath: *Empiricism and Sociology.* Edited by Marie Neurath and R. S. Cohen)
J. W. N. Watkins

352 ORTHODOXY IN QUANTUM MECHANICS (review of E. Scheibe: *The Logic of Quantum Mechanics*)
M. L. G. Redhead

page

358 PROPENSITIES AND PROBABILITIES (review of current literature on the propensity interpretation of probability)
Henry E. Kyberg

375 MAURICE CLAVELIN ON GALILEO'S NATURAL PHILOSOPHY (review of M. Clavelin: *La Philosophie Naturelle de Galilée. Essai sur les Origines et la Formation de la Mécanique Classique*)
Howard Stein

REVIEWS

BAR-HILLEL, Y. (*ed.*) Pragmatics of Natural Language, review by J. L. Mackie, 83

WINOGRAD, T. *Understanding Natural Language*, review by M. A. Boden, 85

WHITROW, M. *ISIS Cumulative Bibliography of the History of Science formed from ISIS Critical Bibliographies* 1–90, 1913–1965, review by J. E. Murdoch, 89

AJDUKIEWICZ, K. *Problems and Theories of Philosophy*, review by J. Giedymin, 189

BARTLEY, W. W. *Wittgenstein*, review by I. C. Jarvie, 195

NELSON, L. *Progress and Regress in Philosophy*, review by A. Mannu, 198

RESCHER, N. *The Coherence Theory of Truth*, review by D. Miller, 291

SIMON, M. A. *The Matter of Life*, review by J. Garner, 295

SKLAIR, L. *Organised Knowledge*, review by N. Koertge, 297

BODEN, M. *Purposive Explanation in Psychology*, review by R. Gregory, 299

MISCELLANEOUS

92–97 RECENT PUBLICATIONS

398–400 RECENT PUBLICATIONS

400 BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE
Programme of Meetings, 1974–1975

401 CENTRE FOR THE PHILOSOPHY OF SCIENCE, UNIVERSITY OF PITTSBURG
Fourteenth lecture series, 1974–75

INDEX OF AUTHORS

ARTICLES

Graves, J. C. Uniformity and Induction, 301

Musgrave, A. E. Logical versus Historical Theories of Confirmation, 1

Post, J. Propositions, Possible Languages and the Liar's Revenge, 223

Puccetti, R. Pattern Recognition in Computers and the Human Brain, 137

Schlesinger, G. The Unpredictability of Free Choices, 209

Urbach, P. Progress and Degeneration the 'I.Q. Debate' (I) and (II), 99 and 235

DISCUSSIONS

Coats, A. W. Situational Determinism in Economics: The Implications of Latsis's Argument for the Historian of Economics, 285

Cohen, L. J. Professor Hull and the Evolution of Science, 334

Feyerabend, P. K. Zahar on Einstein, 25

Good, I. J. A Correction Concerning Complexity, 289

Good, I. J. A Little Learning Can Be Dangerous, 340

Harris, J. H. Popper's Definitions of Verisimilitude, 160

Hull, D. L. Are the 'Members' of Biological Species 'Similar' to Each Other? 332

Machlup, F. Situational Determinism in Economics, 271

Miller, A. I. On Lorentz's Methodology, 29

Miller, D. Popper's Qualitative Theory of Verisimilitude, 166

Miller, D. On the Comparison of False Theories by their Bases, 178

Nartonis, D. K. Quantum Fact and Fiction, 329

Prokhovnik, S. J. Did Einstein's Programme Supersede Lorentz's? 336

Quinn, P. L. The Transitivity of Non-Standard Synchronisms, 78

Redhead, M. L. G. On Neyman's Paradox and the Theory of Statistical Tests, 265

Schaffner, K. F. Einstein versus Lorentz: Research Programmes and the Logic of Theory Evaluation, 45

Scheibe, E. Popper and Quantum Logic, 319

Spielman, S. On the Infirmities of Gillies's Rule, 261

Tichý, P. On Popper's Definitions of Verisimilitude, 155

REVIEW ARTICLES

Kyberg, H. E. Propensities and Probabilities (review of current literature on the propensity interpretation of probability), 358

Redhead, M. L. G. Orthodoxy in Quantum Mechanics (review of E. Scheibe *The Logic of Quantum Mechanics*), 352

Stein, H. Maurice Clavelin on Galileo's Natural Philosophy (review of M. Clavelin *La Philosophie Naturelle de Galilée. Essai sur les Origines et la Formation de la Mécanique Classique*), 375

Watkins, J. W. N. Otto Neurath (review of O. Neurath *Empiricism and Sociology*, 343

REVIEWS

Boden, M. A.	T. Winograd: *Understanding Natural Language*, 85
Garner, J.	M. A. Simon: *The Matter of Life*, 295
Giedymin, J.	K. Adjukiewicz: *Problems and Theories of Philosophy*, 189
Gregory, R.	M. A. Boden, *Purposive Explanation in Psychology*, 299
Jarvie, I. C.	W. W. Bartley: *Wittgenstein*, 195
Koertge, N.	L. Sklair: *Organised Knowledge*, 297
Mackie, J. L.	Y. Bar-Hillel (*ed.*): *Pragmatics of Natural Languages*, 83
Mannu, A.	L. Nelson: *Progress and Regress in Philosophy*, 198
Miller, D.	N. Rescher: *The Coherence Theory of Truth*, 291
Murdoch, J. E.	M. Whitrow, *Isis Cumulative Bibliography*, 89

Brit. J. Phil. Sci. 25 (1974), 99–135 *Printed in Great Britain*

Progress and Degeneration in the 'IQ Debate' (1)*

by PETER URBACH

Introduction

1 *The Hereditarian and Environmentalist Research Programmes.*
  (*a*) *The Hard Cores of the Two Programmes.*
  (*b*) *The Positive Heuristics of the Two Programmes.*
  (*c*) *Should Scientists Believe the Hard Cores of Their Programmes?*
  (*d*) *The Theory of General Intelligence* versus *the Faculty Theory.*

2 *The Progress of the Hereditarian Programme.*
  (*a*) *The Inheritance of Intelligence.*
  (*b*) *Intelligence and Social Class.*
    (*i*) *Filial Regression to the Mean and Social Mobility.*
    (*ii*) *Dysgenic Effects.*
  (*c*) *The Distribution of Intelligence.*
  (*d*) *Racial Differences in Intelligence.*
  (*e*) *Conclusion.*

3 *The Degeneration of the Environmentalist Programme.*
  (*a*) Ad Hoc *Explanations of IQ Differences in Terms of Socio-Economic and Cultural Factors.*
  (*b*) Ad Hoc *Explanations of IQ Differences in Terms of Personality Factors.*
    (*i*) *Motivation, Internal Control, Self-Esteem and Father-Absence.*
    (*ii*) *The Sensory Deprivation Theory.*
    (*iii*) *The Racial Stress Theory.*
  (*c*) *Tests of Environmentalist Theories in Compensatory Education Schemes.*
  (*d*) *The Inheritance of the Environment.*
  (*e*) *Conclusion.*

Received 30 *November* 1973

---

* I owe my greatest intellectual debt to the late Imre Lakatos. Not only was his methodology of scientific research programmes the crucial philosophical influence on this paper, but he also personally devoted an immense amount of time to constructive criticism of earlier drafts.

I also gratefully acknowledge the helpful criticisms of earlier drafts of this paper made by Peter Clark, Colin Howson, John Watkins, Elie Zahar and especially John Worrall, none of whom should, however, be taken as agreeing with all I say.

INTRODUCTION

In the question of intelligence, all roads lead back to the work of Francis Galton in the late nineteenth century. He was the first to formulate a theory of 'general intelligence'[1]; he pioneered the construction of mental tests; and he was the inventor of experimental methods for investigating the inheritance of mental abilities.

The controversy between those who hold that individual and group differences in intelligence test scores are primarily due to inherited differences and those who hold they are primarily the products of environmental differences is one of the oldest and most acrimonious in social science. The controversy has been almost universally marred by a lack of clearly defined standards by which to assess the rival theories. Whenever such standards *have* been invoked, they have been either utopian[2] or unsatisfactory.[3]

I intend to describe and appraise the rival views of intelligence in terms of Lakatos's methodology of scientific research programmes and thereby to evaluate the two competing programmes in the light of their objective merits. This will also detach the debate from the political positions with which they are falsely associated in the public mind.

My thesis is that the hereditarian-environmentalist rivalry has existed not between two competing *theories* but between two competing *research programmes*. For there have been a *series* of—falsifiable and, indeed, falsified—hereditarian and environmentalist theories, each term in each of the two series dealing with some of the refutations (or anomalies) faced by its predecessor. Each of the two series can be characterised by a set of assumptions (*'hard cores'*) common to all the terms. Moreover, each of the series is associated with certain 'heuristic' machinery guiding the construction of successive hypotheses in the series. Thus both the hereditarian and environmentalist approaches possess all the features identified by Lakatos in major physical scientific research programmes.[4]

---

[1] 'General intelligence' is a technical term; *cf. below*, p. 111 and section 1 (*d*).

[2] *Cf. e.g.* Deutsch ([1969], p. 527), who uses inductivist standards and complains that the hereditarian theory goes 'far beyond what is warranted by the data'. But *all* theories go far beyond the 'warranted data': *cf. e.g.* Popper [1934], section 25.

[3] *Cf. e.g.* Burt and Howard ([1957], p. 55), who use probabilist standards and claim that 'the only way to rebut this [the hereditarian] argument is for the critic to produce a constructive hypothesis of his own, *and formally demonstrate its higher probability*' (my italics). But despite many efforts, no satisfactory way has been found (and none is likely to be found) to rank scientific theories according to their probability: *cf.* especially Lakatos [1968*a*]. *Cf.* also Burt [1957], p. 164, for a defence of hereditarianism on grounds of a 'consilience of inductions'.

[4] For the methodology of scientific research programmes, *cf.* primarily Lakatos [1970], [1971*a*] and [1971*b*]; but *cf.* also his [1963–4], the first case study of a research programme (on polyhedra).

In section 1 I shall *describe* the rival hereditarian and environmentalist research programmes. In sections 2 and 3 I shall *appraise* the development of the two programmes.

As Lakatos has shown—referring to the conventionalist arguments of Le Roy and Milhaud[1]—it is relatively easy for any programme to deal with (that is, make its theories consistent with) any given anomalies. In appraising the programmes, the question is whether they do this in a *progressive* or in a *degenerating* manner. The shift within a research programme from one theory to another is progressive if the new theory not only deals with its predecessor's anomalies but also makes extra predictions, some of which are tested and confirmed. On the other hand, if the new theory does nothing more than accommodate the anomalies, the shift is *ad hoc* and the programme degenerating.

The question we must ask about the hereditarian and environmentalist programmes is *not* whether they have been proved from the facts (no programme could be), nor whether they are refuted (even the latest versions of both programmes *are* refuted). On Lakatos's criterion, the fact that a programme is refuted does not, by itself, render it unscientific or unacceptable. The question is whether either programme has successfully 'stuck its neck out' by predicting experimental results which had not been known in advance or whether it simply tinkers with its assumptions in such a way as to accommodate the already observed facts without successfully anticipating further, yet unobserved, *novel* facts, in the general spirit of the heuristic of the research programme.[2]

1 THE HEREDITARIAN AND ENVIRONMENTALIST RESEARCH PROGRAMMES

(*a*) *The Hard Cores of the Two Programmes.*

The hard core of any research programme is not in general likely to be recognisable from the declarations of the scientists involved since scientists do not necessarily *believe* in the hard cores of the programmes on which they work and they sometimes deny them.[3] Moreover, a programme's hard core might never have been asserted or recognised. However, although the private beliefs of scientists provide no guide to the hard core of their programme, this may be identified from the principles which direct their practical research. For the hard core of any research programme has a prescriptive counterpart[1]; that is, it is intimately associated with a certain characteristic ***heuristic machinery*** which provides strategies for the recognition and resolution of anomalous experimental results.[2]

---

[1] *Cf.* Le Roy [1899] and Milhaud [1896].

[2] While this paper was being written, Zahar introduced an interesting modification of the concept of 'novel fact' (Zahar [1973], especially pp. 101–4) when analysing the Einsteinian Revolution; and the new standard of appraisal was also successfully applied to the Copernican Revolution (Lakatos and Zahar [1975]). However, the application of Zahar's modification did not lead to any alteration in the conclusions of the present paper.

[3] For some examples of scientists who have disavowed the hard core of the programme on which they worked, *cf. below*, section 1 (*c*).

In this paper, I shall argue that the rivalry in the 'IQ Debate' exists, not between scientists holding this or that *belief*, but between scientists carrying out their research according to the prescriptive rules contained in one or other of two separate heuristics. These are the hereditarian and environmentalist heuristics which I shall describe *below*.[3] I have also connected these two heuristics with corresponding hereditarian and environmentalist hard cores since, if science is more than a mere game, the heuristics must be attached to some propositions describing the world.

Now some scientists may not believe the hard cores I have attributed to their research programmes and others may even disavow them altogether but the fact that their research is performed in accordance with the corresponding heuristic principles entitles us to say that these scientists work *as if* they were inspired by them.

The ***hard core of the hereditarian programme*** consists of two propositions:

(1) ***All individuals possess a general mental capacity called 'general intelligence' which enters with some (and varying) degree into all the diverse types of cognitive activity.***[4]

(2) ***Differences between individuals and between groups in 'general intelligence' are the results of inherited differences.***

The ***hard core of the environmentalist programme*** consists of the proposition:

(1) ***All individuals inherit all sorts of different mental capacities but each is identical in each individual; all differences in cognitive abilities are the results of environmental differences.***

Although there is a broad environmentalist programme which covers all mental capacities, I shall in this paper consider only the more specialised version of this research programme which concerns the general mental capacity ('general intelligence') whose existence is postulated in the first proposition of the hereditarian hard core.

Some people think that the 'IQ Debate' is more accurately represented by *one single*, 'interactionist', research programme with a hard core somewhere intermediate between the hereditarian and environmentalist hard cores I have described. I shall consider this view in detail and show why it is mistaken later.[1] Readers will be able to judge whether research on intelligence truly constitutes *two* research programmes with different hard cores and different heuristics when I come to describe the actual research in section 2 and in section 3 which will appear as Part II in the next issue of this *Journal*.[2]

---

[1] For the connection between metaphysical hard cores and heuristics, *cf.* Watkins [1958].

[2] 'Hard core' and 'positive heuristic' are technical terms of the methodology of scientific research programmes: *cf. e.g.* Lakatos [1970], pp. 132–8.

[3] *Cf. below*, section 1 (*b*).

[4] This at first sight surprising claim will be explained and discussed in section 1(*d*). As I shall show there, the claim and the central notion of *general* intelligence were sharpened in a battle with the rival theory—the so-called 'faculty theory'.

Logically the hard cores of both programmes are all-some statements and they consequently have no potential falsifiers.[3] Falsifiable versions can only be generated by conjoining them with some auxiliary, 'protective belt', hypotheses.

Thus the hereditarian hard core in itself has no empirical content and it only gains predictive power when it is embedded in a particular theory of inheritance. Galton and his student Karl Pearson held to the blending theory of inheritance[4] but at least since Fisher's seminal studies around

---

[1] *Cf. below*, pp. 110–11.

[2] Although, as I shall argue in section 1(*c*), the private beliefs of scientists are not relevant to an objective description and appraisal of their research, some researchers in the field of intelligence have articulated the hard cores of the programmes on which they work.

Consider first the hereditarians. For example, Galton expresses himself uncompromisingly: 'I have no patience with the hypothesis occasionally expressed and often implied, especially in tales written to teach children to be good, that babies are born pretty much alike, and that the sole agencies in creating differences . . . are steady application and moral effort. It is in the most unqualified manner that I object to pretensions of natural equality' (Galton [1869], p. 14). Cyril Burt characterises intelligence as follows: 'First, it is a *general* quality; it enters into every form of mental activity; secondly, it is (in a broad sense of the word) an intellectual quality—that is, it characterises the cognitive rather than the affective or conative aspects of conscious behaviour; thirdly, it is inherited or at least *innate*; differences in its strength or amount are due to differences in the individual's genetic constitution' (Burt [1955], p. 163). Burt also conjectures that 'most psychologists believe that differences in intelligence are innate . . . because of the vast mass of converging evidence' (Burt [1943], p. 89).

Let us now cite some environmentalists. One of the earliest explicit statements of the main tenet of environmentalism was made by Descartes. He writes: 'the power of forming a good judgment and of distinguishing the true from the false, which is properly speaking what is called Good sense or Reason, *is by nature equal in all men*'. He adds modestly that 'I have never ventured to presume that my mind was *in any way* more perfect than that of the ordinary man' (Descartes [1637], p. 82; my italics). More recently, J. B. Watson ([1931], p. 270) claimed that 'all healthy individuals . . . start out *equal*' and he also declared: 'Give me a dozen healthy infants, well-formed, and my own specified world to bring them up in, and I'll guarantee to take any one at random and train him to become any type of specialist I might select—doctor, lawyer, artist, merchant-chief, and, yes, even beggar-man or thief, regardless of his talents, penchants, tendencies, abilities, vocation, *and race of his ancestors*' (*ibid.* p. 104; my italics). In a U.N.E.S.C.O. statement of 1950, a group of sociologists and anthropologists assert that 'given similar degrees of cultural opportunity to realise their potentialities, the average achievement of members of each ethnic group *is about the same*. The scientific evidence of recent years *fully support* the dictum of Confucius . . . [that] "men's natures are alike; it is their habits that carry them far apart"' (U.N.E.S.C.O. [1950]; my italics).

[3] *Cf. e.g.* Watkins [1958], pp. 345–8; but *cf.* Lakatos [1970], pp. 144–5.

[4] According to the blending theory of inheritance, the characters of a child are always a blend of the characters of its parents. This theory implies that successive generations would display diminishing variation in all inherited characters, a prediction which was

1918[1] the hereditarian programme has assumed that intelligence is inherited by the multifactorial mechanism. The multifactorial theory of inheritance asserts that characters showing continuous quantitative variation are determined by a large number of genes which segregate in accordance with Mendelian principles and which each produce a small, similar and cumulative effect.

Even when conjoined with the multifactorial theory, the hereditarian programme's hard core makes no claims about the relative intelligence of any particular pair of individuals (unless they happen to be monozygotic twins) or of any two racial or social groups. Instead, the programme makes *statistical* predictions concerning the values of correlations for intelligence between large numbers of pairs of variously related persons.

On the other hand, the environmentalist programme is much 'bolder', in Popper's sense. Unlike its rival, it does not lead merely to statistical predictions but asserts that *every pair* of individuals inherits identical innate mental capacities. Thus, the environmentalist programme enables us to make predictions concerning phenomena about which the hereditarian programme has not been able to make any. For example, the environmentalist programme anticipates that all social and racial groups will have identical grades of intelligence while the hereditarian programme is completely agnostic about the relative average intelligence of different groups.

Now both research programmes make predictions concerning inherited mental capacities and in order that these predictions may be tested, it is necessary that instruments be devised which will measure these capacities. Such instruments have become known as IQ tests.

In assessing the methodological role of IQ tests in each of the research programmes, thermometers provide an instructive analogy, for the relationship of thermometers to thermodynamics is rather similar to that of IQ tests to theories of intelligence. Since thermodynamics is a theory about heat, it requires thermometers in order that its predictions may be empirically tested. However, the development of reliable thermometers took several centuries and even today we do not have any *perfect* thermometers. Particular difficulty was encountered when trying to devise thermometers which would accurately measure extremes of temperature. Under such extreme conditions, some auxiliary assumptions on which conventional thermometers were based, such as the linear proportionality between

---

controverted by the facts. Darwin tried to resolve the anomaly by postulating the spontaneous development of new variations. But this was a sadly *ad hoc* move and the anomaly was not resolved progressively until the rediscovery of Mendel's theory around 1900.

[1] Fisher [1918].

change in heat and the volume of the thermometer fluid, had to be abandoned. Such auxiliary assumptions form part of the 'protective belt' of the thermodynamics programme and they are accepted if they are 'progressive', that is independently testable and corroborated, otherwise they may be discarded.

In constructing IQ tests, various auxiliary assumptions are made. For example, it is assumed that the cultural content of the questions in the test is equally familiar and significant for all potential subjects. Each of the programmes then uses the resulting IQ tests to check its respective predictions concerning general intelligence. If any predictions are not confirmed, then researchers frequently replace some auxiliary assumption which had been taken for granted when the test was constructed and in this way they try to improve their tests. Such techniques constitute the *heuristic* of the two programmes.[1]

It is clear now why I put 'IQ debate' in inverted commas in the title of my paper. To call the controversy the 'IQ debate' is like calling the rivalry between two theories of heat the 'thermometer debate'. The misnomer is due to the strong operationalist influence among scientists, some of whom believe, falsely, that science is about what measuring instruments measure. But acceptance of operationalism precludes any explanation of how one measuring device is better than another.[2]

(*b*) *The Positive Heuristics of the Two Programmes.*

Each of the programmes has developed manoeuvres which are employed to resolve anomalous test results and which guide the construction of improved IQ tests.

The *heuristic of the hereditarian programme* has developed considerably since its main tenets were set out by Galton. In its present form it contains R. A. Fisher's techniques for analysing variance, the machinery of factor analysis developed by Spearman, Burt, Thurstone and others and at least two methodological directives.

First, the heuristic directs hereditarians to construct ever-improving tests of 'general intelligence' and to check these tests by using them to measure the IQ's of people whose genetic relations are known from Mendelian theory.[3] If these tests do not yield the predicted pattern, then the hereditarian first blames the test. The test is declared 'badly administered'. For example, it may be conjectured that the subjects were not put sufficiently at ease during the test.[1] If no such assumption succeeds in dissolving the anomaly, the test is declared 'culture biased' against some people, in other words it is said to favour those people who possess some specialised knowledge or experience. In order to test this assumption, a new test must be employed (or specially constructed) which can be seen to exclude the putative unfair cultural element. One can, of course, introduce more substantial changes in order to account for anomalies.

---

[1] *Cf. below*, section 1(*b*).

[2] For a good criticism of operationalism, *cf.* Zahar [1973].

[3] The hereditarian heuristic originated with Galton in whose work we find the first attempts to measure intelligence. In these pioneering attempts, the appraisal of intelligence involved judging a person's 'eminence' by surveying his biography. (*Cf. below*, p. 115, footnote 1.) In order to show that differences in intelligence arose from inherited differences, Galton, guided by the blending theory of inheritance, tried to demonstrate

For instance, one may change the underlying theory of inheritance as applied to intelligence.[2] (One might even invent a radically new theory of intelligence and replace the most basic ingredients of the background knowledge which had supported the research programme. A revolution in gene physiology may well make us discard both rival research programmes, but such major creative innovations will hardly be triggered off by mere anomalies in the relatively small hereditarian programme. It is however logically possible for hereditarians to dissolve anomalies by appealing to profound environmental effects which influence the genes or which hamper their normal action. Such anomaly-resolving techniques are not employed in the hereditarian programme. No doubt, this is simply a result of the fact that no satisfactory theories describing the chemistry and physics of 'intelligence genes' are available. When such theories come forward, the present day controversies are likely to take radically new forms.)

New tests are generally constructed, not in response to anomalies but in order to extend their scope, for example, to entirely different cultures or to handicapped children. The most ambitious efforts along these lines have involved attempts to construct 'ideal' IQ tests which are completely

---

that the closer two people are related, the more similar they are in qualities of 'eminence'. He also invented experimental methods in an attempt to discriminate between hereditarian and environmentalist theories current in his day. For example, he was the first to propose the study of twins and of adopted children. (*Cf.* Galton [1883] and [1889].) He was aware that such crude and subjective assessments of intelligence were, at best, first approximations. The development of objective measures of native intelligence was a natural part of his research programme. For Galton, the aim of objective mental tests was 'to obtain a general knowledge of the capacities of man by sinking shafts, as it were, at a few critical points' (Cattell [1890], footnote inserted by Galton).

[1] *Cf. e.g.* Jensen's explanation of some alleged increases in children's IQ's in programmes for compensatory education, *cf.* below, section 3 (*c*) (in Part II).

[2] This course was adopted by Karl Pearson in the early days of the hereditarian programme when he reverted to the blending theory of inheritance after his sibling correlations for 'ability' and stature failed to conform to the predictions of the multifactorial theory. This was an *ad hoc* manoeuvre. The failure of IQ's to follow a normal curve over the whole range of IQ's also led some researchers to modify slightly the multifactorial theory of inheritance as applied to intelligence. This manoeuvre was not *ad hoc*. (*Cf. below*, pp. 128–30.)

culture fair, in so far as they require no knowledge of any particular culture but depend on physiological responses.[1]

There is a second methodological directive in the hereditarian heuristic, namely to compare the intelligence of different groups and to investigate the relationship between intelligence and important social phenomena.[2]

Comparisons of the intelligence of different groups may throw light on some important sociological phenomena since if two ethnically similar but geographically or socially displaced groups are found to have different average IQ's, hereditarians (if they do not invoke the culture bias of the tests) may adopt the auxiliary assumption that one of the groups was derived from the genetically more intelligent members of the original group by *selective migration*.[3]

The *heuristic of the environmentalist programme* contains the directive to search for the *environmental* factors which bring about differences from original intellectual equality. Whenever a test of intelligence, which is initially assumed to be 'culture fair' yields individual or group differences, environmentalists account for these differences by conjecturing that the 'less able' subjects had been exposed to one or more (previously hidden) deleterious experience or had been deprived of some (previously unrecognised) environmental stimulus. (The effect which the environmental factor has on a particular individual may be reversible or irreversible.) In order to test such conjectures, the environmentalist may either repeat the same IQ test on individuals who are equal in regard to the *specified* factor or he may assume that the test was culture biased after all and change the test material so that the guilty environmental factor is excluded from it. The conjecture will be confirmed if the individuals now gain more nearly equal scores. By identifying specific environmental factors which affect performance on intellectual tasks and by compensating for them, the environmentalist programme aims to construct tests on which all environmentally equal subjects gain the same scores.

---

[1] Galton and Cattell made the earliest such attempts by trying to measure intelligence *via* the sensory motor functions. (*Cf.* Cattell [1890].) These efforts were, however, abruptly abandoned when it was discovered by Wissler (*cf.* his [1901]), using the method of correlation newly introduced by Pearson, that no correlation at all existed between such measures and intuitive notions of intelligence. (The importance of intuitive notions of intelligence as points of departure for both programmes is discussed *below*, p. 108.) More recently, researchers have had some success along these lines by finding very high correlations between the results of standard IQ tests and certain quantities connected with the 'evoked potential' produced when the brain receives a sudden physical stimulus. (For a review of these investigations, *cf.* Eysenck [1973], pp. 78–81.)

[2] Although the 'IQ debate' has tended, at least in popular expositions, to centre round the issue of social class and racial differences, many other interesting social phenomena come within the purview of the two programmes. For example, a possible link between criminality and intelligence could be investigated. Such studies were envisaged by Binet in the earliest days of mental testing. *Cf.* Binet and Simon [1905*b*], p. 246 (quoted in Wolf [1969], p. 215).

[3] *Cf. e.g.* the hereditarian explanation for the origins of social class differences in average IQ score, *below*. pp. 122–5.

From my description of the heuristics of the two programmes it is apparent that there is considerable (but by no means complete) overlap between them. The heuristics differ most importantly in that, because of their different hard cores, they make different predictions concerning the distribution of intelligence and hence they recognise different events as anomalies. The two programmes, however, share the common aim to construct tests of general intelligence which tap *inborn capacities* rather than acquired skills. Hereditarians and environmentalists share crude and intuitive ideas about the nature of cognitive capacity, in contradistinction to acquired skills and in contrast with emotional reactions. Various intuitive notions of intelligence have been articulated by environmentalist and hereditarian researchers. For example, intelligence has been characterised as 'good sense or reason' (Descartes [1637]), the ability 'to judge well, to understand well, [and] to reason well' (Binet 1905*a*]), 'the ability to educe relations and correlates' (Spearman [1927]), 'the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment' (Wechsler [1944]), and as 'problem-solving ability' (Davis [1949]).

Some critics allege that the fact that researchers have been unable to characterise intelligence more precisely constitutes a fatal weakness in their programmes. But they are wrong. It is common in science for initially vague concepts to attain an improved and more precise formulation by being embedded in a growing research programme. For instance, heat, mass and force, once primitive, undefined ideas, were modified and given precision within developing thermodynamics and dynamics. Just as 'intelligence' is sometimes derided as being 'meaningless',[1] the concept of force acting at a distance was regarded by some seventeenth-century critics as 'occult'. In appraising whether the initially intuitive concepts of intelligence are *sterile* or *fertile*, it will be necessary to see whether they have been incorporated into degenerating or progressive research programmes. This task of appraisal will be carried out in sections 2 and 3.

(*c*) *Should Scientists Believe the Hard Cores of Their Programmes?*

It will no doubt be objected that, in setting up rival hereditarian and environmentalist positions, I have grossly misrepresented the debate since few of the scientific protagonists actually *believe* either of the hard cores of the two research programmes. Now it is certainly the case that many environmentalists and hereditarians would not declare belief in the hard cores of their respective programmes, although some do.[1] But the private, psychological beliefs of scientists have no place in an appraisal of their work. What must be considered in describing their research are the heuristic principles which they follow *in practice*. It is perfectly possible for a scientist to work within a research tradition without *believing* that the assertions about the world on which the tradition is based accurately reflect reality. Take P. E. Vernon, for example. He declares that he is 'not sympathetic to the view that intelligence itself is an innate quality and that our tests are inefficient at measuring it because they are subject to environmental influences'.[2] Yet in the same paper from which this quotation is taken, Vernon attempts to show how *all* of the alleged increases in the IQ's of some 'feeble-minded' children reared in special environments can be traced to the inadequacies of the tests themselves or to the way in which they were administered. Caldwell is an example of a researcher who follows the heuristic of the environmentalist programme despite the fact that she 'pays lip service' to the existence of genetic differences in intelligence. She says: 'most of us in enrichment efforts [that is, attempts to increase IQ's by compensatory education[3]]—no matter how much *lip service* we pay to the genetic potential of the child—are *passionate believers* in the plasticity of the human organism. We desperately need to believe that we are born equalisable. With any failure to demonstrate the effectiveness of compensatory experiences offered to children of any given age, *one is entitled to conclude parsimoniously that perhaps the enrichment was not offered at the proper time*.'[4] Lakatos mentions Newton and Planck as examples of physical scientists who worked on research programmes whose hard cores they did not believe in[5]; another important example is Maxwell in his early work on kinetic theory.[6] It is certainly true that the research in the field of intelligence can be divided into that which is carried out according to the hereditarian and environmentalist heuristics as I have described them. The only way I can see of attaching these heuristics to claims about the world is by associating them with assertions equivalent to those I have characterised as comprising the respective hard cores. It has, however, to be emphasised strongly that the progress and degeneration of a research programme are logically independent of the scientist's degree of belief in or commitment to its hard core. I conjecture that, *at least in the past, the extent to which scientists' psychological beliefs used to approximate to the hard cores of their own programmes is positively correlated with how objectively progressive these programmes were*. A classical corroboration of this conjecture is the early scepticism of Newtonians concerning their principle of 'action at a distance' in contrast with the dogmatism of later Newtonians. Another example is the scepticism of early Copernicans, like Osiander, towards the hard core of their programme, and their increasing belief in it as its progress became overwhelming.[1]

---

[1] *Cf. e.g.* Maddox [1957].

---

[1] *Cf. above*, p. 103, footnote 2.

[2] Vernon [1951], p. 136.

[3] *Cf. below*, section 3 (*c*) (in Part II).

[4] Caldwell [1968], p. 81; my italics.

[5] Lakatos [1970], p. 144–5.

[6] *Cf.* Brush [1965], pp. 26[illegible]

Before proceeding from the *description* of the two rival programmes to their *evaluation*, I would like to counter one standard counterargument. According to some psychologists, research on intelligence must be represented by *one single research programme* and not by *two* rival programmes. Eysenck, who puts this view most clearly, suggests that there is *only one single programme* which he calls the '*interactionist*' programme and whose main tenet is that 'genetic factors, interacting with environmental ones are active in producing the observed . . . differences'[2] and the task of the interactionist is 'to *find out* the relative importance of these factors'.[3] Indeed, Eysenck argues for the much stronger position that 'interactionism' is 'the only tenable view'[4] since, he maintains, 'a purely hereditarian . . . [theory] which would account for *all* the observed phenomena [that is, IQ differences] in terms of genetic factors alone . . . would run foul of the most elementary tenets of genetic science, with its stress on the difference between genotype and phenotype'.[5] But although Eysenck insists that the *possibility* that all IQ differences are caused by genetic differences is ruled out on theoretical grounds, he does not specify exactly what percentage of IQ differences we may legitimately consider to be genetic in origin without violating 'the tenets of genetic science'. Nor could he. The truism that there can be no phenotype unless the genotype comes into contact with an environment does not preclude the *possibility* that the differences in a particular phenotype might be exclusively the result of differences in genotype.

Now Eysenck's 'interactionism', with its objective of 'finding out' the relative importance of hereditary and environmental factors in producing IQ differences is not a research programme at all, any more than an exercise in 'finding out' the chemical composition of a particular sample of air would constitute a research programme. Eysenck's view may accurately describe the checking of *a particular intelligence test* but it overlooks the vital fact that hereditarians, for example, do not meekly acknowledge that there are environmental influences on intelligence when evidence for such influences is presented but, guided by the heuristic of their programme, they usually invent some auxiliary hypothesis which makes the anomalous result consistent with hereditarian theories. The task of the methodologist is to assess whether these auxiliary hypotheses are *ad hoc* or not and hence whether the programme in which they figure is progressive or degenerating.

---

[1] Lakatos predicted that one major future branch of psychology will concern itself with causal relations between the world of objective ideas (Frege's and Popper's 'third world' of ideas; *cf.* Popper [1972], chs. 5 and 6) and psychological and even moral factors. For instance he conjectured that protagonists of degenerating research programmes tend to succumb to moral degeneration. (Private communication.) This is a more specific version of a prediction made by Popper in his [1972] (p. 156).

[2] Eysenck [1971*a*], p. 115.

[3] *Ibid.*; my italics.

[4] Eysenck [1973], p. 29.

[5] Eysenck [1971*a*], p. 115.; my italics.

There is, however, some truth-content to Eysenck's thesis, namely that all research programmes are anomaly-ridden, all hard cores are likely to be false and normally each of any two rival programmes has a grain of truth in it. But the methodology of scientific research programmes is basically conventionalist and hence fallibilist and while it recognises the impossibility of comparing the *truth-content* of rival theories it claims to be able to appraise the empirical success of different problem-recognising and problem-solving machinery.

(*d*) *The Theory of General Intelligence* versus *the Faculty Theory.*

Galton was the first to propose that individuals possess a *general* intellectual ability which threads the whole range of their mental abilities, as well as some special aptitudes of lower generality. According to Galton's theory, people can be arranged in a '[*linear*] continuity of natural ability'.[1] He adduced some anecdotal evidence to show that great men of science owe their eminence much less to special powers than to the concentration of their general giftedness in some particular direction.[2]

Galton's theory was in sharp contrast with the rival *faculty theory* of the intellect which was especially championed in this century by Thorndike, Kelley and Thurstone.[3] The faculty theory *denies* the existence of a general intellectual factor and asserts instead that the intellect is composed of many *independent faculties* such as mathematical, mechanical and verbal faculties. The faculty theory predicts that people can only be arranged multi-dimensionally into qualitatively different types possessing different strengths of different faculties.

The two theories were made testable by Charles Spearman and others who imported precise statistical techniques into the debate. Spearman pointed out that if general intelligence exists, then all tests of mental ability should be positively correlated.[4]

Indeed, Spearman's suggestion was carried out and the observations favoured the theory of general intelligence. When the scores gained by people on a large number of diverse mental tests are intercorrelated, the correlations are pair-wise positive as predicted by the theory of general intelligence. For example, high correlations exist between such apparently diverse activities as vocabulary tests, discovering the shortest way out of a printed maze, memorising digits and copying designs.[1]

---

[1] Galton [1869], p. 26.

[2] *Ibid.*, p. 22.

[3] *Cf.* Thorndike [1925], Thurstone [1938] and Kelley [1928].

[4] Spearman [1904].

Thurstone, however, attempted to demonstrate in his [1938] that this is not so and that evidence corroborated the faculty theory. He devised 56 tests of 'primary mental abilities' and analysed the scores which 240 students gained on them. Using a refinement of the method of factor analysis, developed by Spearman, Thurstone extracted eight apparently independent 'primary mental factors' which he characterised as verbal, perceptual speed, inductive reasoning, word fluency, number, rote memory, deductive reasoning and spatial abilities.

The apparent absence of a general factor common to all of Thurstone's tests was a corroboration for the faculty theory, but the victory was short-lived. Within months, Spearman and Eysenck had 'reworked' Thurstone's data and demonstrated that a general factor did exist and had been suppressed merely by the application of a specially chosen statistical procedure.[2] The main reason why the general factor could be so easily concealed in Thurstone's results was that he used subjects whose range of ability was very narrow. On repeating the experiment with children representing a wider spectrum of ability this general factor was much more prominent and could no longer be eliminated by the statistical methods Thurstone had originally employed.[3] Although Thurstone played this general factor down and dubbed it a 'second order factor',[4] most contemporary researchers (whether hereditarian or environmentalist) now accept the theory of *general* intelligence.

The principles used in attempts to construct intelligence tests are dictated by the theory of intelligence adopted. This may be illustrated by comparing some of the *desiderata* which the faculty theory and the theory of general intelligence impose on intelligence tests.

---

[1] Actually, Spearman advocated the much stronger thesis than the theory of general intelligence I have described. This is the so-called 'two factor theory' according to which an individual's performance on all *diverse* mental tests can be divided into a general factor (general intelligence) threading *all* tests of intellectual ability and a second factor *specific to each separate test*. Thus Spearman denied the existence of any special factors of lower generality than general intelligence which are common to sub-groups of diverse mental tests. (*Cf. e.g.* Spearman [1904].) When evidence for such special factors was discovered among such diverse tests, Spearman resorted to the sadly *ad hoc* assertion that the tests were not *really* diverse. (Spearman [1927].) Thus Spearman's strong theory has degenerated. However, as I have shown, the weaker version of his theory, which postulates the existence of general intelligence, has progressed.

[2] Spearman [1939] and Eysenck [1939].

[3] Thurstone and Thurstone [1941]. *Cf.* also Eysenck [1973], pp. 52–6, for a discussion of Thurstone's programme.

[4] Thurstone [1948].

According to the faculty theory, human aptitudes consist of several independent faculties and intelligence is some kind of combination of these. The ideal intelligence test should consequently contain test items which (*a*) correlate highly with an external criterion of intelligence and (*b*) correlate negligibly amongst themselves. The latter consideration is necessary because if two tests are correlated then they are measuring the same entity to some extent, and to that extent they are repetitive. However, attempts to find test items which satisfy these two *desiderata* have been unsuccessful[1]: the faculty theory has degenerated.

According to the theory of general intelligence, there are quite different *desiderata* on intelligence tests. According to this theory, an intelligence test should contain a large number of 'diverse' items covering as many different subject matters and using as many different types of test material as possible. In addition, the correlations of the items with one another should be high. These two *desiderata* for suitable test items ensure that when the scores gained on each item are added up, the effects of special aptitudes and specific factors which each of them inevitably involves is minimised and the effect of the general factor is maximised.[2] Some attempts to construct tests which satisfy the *desiderata* imposed by the theory of general intelligence were successful, indicating progress in the programme.[3]

---

[1] The psychologists involved in constructing the Army Alpha test tried to construct it from the point of view of the faculty theory but they failed and they fell back on the theory of general intelligence. *Cf.* Yerkes [1921], pp. 316 and 338, and also Freeman [1939], pp. 250–3.

[2] In addition there are various practical considerations which are taken into account when constructing intelligence tests. The most important of these are that the tests are susceptible to easy and objective marking. It is also desirable that the test be available in alternative but equivalent forms so that the effect of coaching and different conditions on the test scores can be investigated. Also standard procedures for administering the test must be worked out. *Cf. e.g.* Terman and Merrill [1937] and Warburton [1969].

[3] The correlation between the scores which a group of people gains on mental tests a and b, r_{ab}, is related to the correlations r_{ag} and r_{bg} which these tests have with hypothetical tests of general intelligence (g) by the relation

$$r_{ab} = r_{ag} \times r_{bg} \qquad (1)$$

The hypothetical correlation r_{ag} is called the 'loading' of test a on general intelligence; or its g-loading. The g-loadings of n mental tests are found by a trial and error method from the $n(n-1)$ observed intercorrelations of these n tests with each other. (*Cf. e.g.* Vernon [1950].) The degree to which an IQ test measures differences in general intelligence is often estimated by averaging the g-loadings of its constituent tests; using this procedure, most IQ tests are found to have g-loadings in excess of 0.7. But this procedure of averaging the g-loadings of the constituent tests surely underestimates the loading of IQ tests on general intelligence; it takes no account of the prediction that combining the scores which people gain on many diverse tests maximises the factor common to them all, namely g, and minimises factors specific to sub-groups of tests. A more satisfactory estimate of the g-loadings of IQ tests may be obtained by using relation 1 in the following manner: the g-loading, r_{vg}, of the vocabulary test in the Stanford–Binet IQ test is about 0.8 and the correlation, r_{vi}, of the vocabulary test with the IQ test of

Note. The first successful intelligence tests were developed by Binet and Simon in France in 1905. As early as 1895, Binet had criticised Galton's and Cattell's attempts to measure intelligence *via* the sensory motor functions. Binet claimed that while elementary psychological processes such as reaction times can be measured with considerable accuracy, intelligence is more likely to be related to higher mental functions, where individual differences are larger.[1] Binet's early experiments were with tests such as those for memory of designs, making up sentences to include specific words, and indicating how two concepts were alike and how they differed. His major criteria for a good test of mental ability were that it should distinguish bright from dull children and that older children should find it easier than younger ones. In addition, he was guided in the selection of test items by his view of intelligence as the quality which enables a person 'to judge well, to understand well, [and] to reason well'.[2]

In 1904, Binet was appointed to a special commission whose task was to recommend methods for the early diagnosis of sub-normal children so that they could be allocated to special schools. In his attempt to find a satisfactory diagnostic test, Binet made the discovery that the scores gained by children on a *combination* of several tests correlated with age and with primitively judged intelligence much better than did each individual test.[3] This was a major empirical success for the theory of general intelligence.

For Binet, a child's intelligence was given by the difference between his mental age and his chronological age. Mental age is defined as follows: when a child's performance on an IQ test and the performance of an average child in a particular age group are similar, the child has a mental age equal to the chronological age of that group. The intelligence quotient, or IQ, which was introduced into the 1916 Stanford revision of the Binet scale[4] is the ratio: $\frac{\text{mental age}}{\text{chronological age}}$. IQ has the advantage over mental age as a measure of intelligence that a particular IQ value represents the same degree of mental retardation at every age. But since mental age levels out at about the age of 15, the ratio of mental and chronological ages is not a suitable measure of intelligence for adults. IQ is now usually given by the ratio of the raw score gained by an individual on a test and the score gained by an average individual of the same chronological age. The tests are standardised so that at each age level, the scores gained by a representative sample of people are approximately normally distributed with a standard deviation of 15 and an average score of 100.

---

which it is a constituent is also about 0.8. By relation (1), the *g*-loading, r_{ig}, of the Stanford–Binet test is about unity. (Strictly speaking, the correlation of the vocabulary test with the score on the *remaining* part of the IQ test should be used in relation (1). But since the vocabulary test contributes as little as 5 per cent to the score on the overall test, the more rigorous approach would produce no significant difference. *Cf.* McNemar [1942], pp. 138–9.)

[1] Binet and Henri [1895].

[2] Binet and Simon [1905*a*], p. 196.

[3] For a discussion of Binet's achievement see Wolf [1969].

[4] The Binet scale was revised by Terman and his co-workers at Stanford University in 1916, 1937 and 1960. The revised scales became known as the Stanford–Binet intelligence tests. *Cf.* Terman [1916] and Terman and Merrill [1937] and [1960].

2 THE PROGRESS OF THE HEREDITARIAN PROGRAMME

(*a*) *The Inheritance of Intelligence.*

The hereditarian programme predicts that general intelligence is inherited. Thus it predicts that there is a positive correlation for intelligence among members of the same family, the size of the correlation being greater the closer the kinship.[1] In particular, the hereditarian programme makes specific predictions concerning the magnitudes of these correlations.

The values predicted by the hereditarian programme for correlations in intelligence of variously related individuals are calculated on the basis of the multifactorial theory of inheritance using statistical techniques largely due to Fisher.[2] The multifactorial theory of inheritance is background knowledge for the hereditarian programme. It asserts that inherited characters showing continuous quantitative variation are determined by a large number of genes which segregate in accordance with Mendelian principles and which each produce a small, similar and cumulative effect. Formulae giving the required correlation coefficients can be derived.[3] Each of these formulae is a function of two parameters which correspond to the effects of partial dominance and assortative mating. The parameter for assortative mating is the correlation for intelligence between parents and it may be directly measured. The parameter giving the degree of dominance may be read off from one of the correlations actually observed or it may be chosen to give the best agreement between all of the observed and theoretical correlations. A consequence of this procedure is that one of the correlations for intelligence cannot be counted as being a prediction.

---

[1] Galton pioneered studies of family similarities in mental qualities. In his most celebrated study, Galton showed that the 100 'most eminent' members of 100 families had 'eminent' relatives in the following numbers: 31 fathers, 41 brothers, 48 sons, 17 grandfathers, 18 uncles, 22 nephews, 14 grandsons, 3 great-grandfathers, 5 great-uncles, 13 first cousins, 10 great-nephews, 3 great-grandsons and 31 other more remote relatives. Galton sought to show that his results were inconsistent with a purely environmental interpretation by pointing out the existence of 'two classes of men with equal advantages in one of which they have high hereditary gifts while in the other they have not'. He also pointed out that current environmental theories could not account for the fact that many brilliant men, such as Newton, D'Alembert, Gauss, Laplace, Whewell, Mill, Faraday, Kant and Kepler were the sons of poor peasants or artisans (Galton [1869]).

[2] While early Mendelian theory could explain qualitative, discrete differences in terms of single gene pairs, it could not account for the types of continuous variation encountered by biometricians. Karl Pearson, a pupil of Galton's found a correlation of about 0.50 between pairs of siblings for various physical traits such as eye colour, height, colour of hair and cephalic index and a similar correlation for 'psychical' characters such as teachers' ratings of 'vivacity' and ability. Pearson argued that since the physical and psychical traits gave similar correlations for siblings they are both inherited. The observed value of 0.50 was however much higher than the theoretical value calculated on the basis of Mendelian theory. Pearson consequently *rejected* Mendelian theory and fell back on the old blending theory of inheritance. (*Cf.* Pearson [1904].) The inconsistency between the measured correlations for graded characters and those predicted by Mendelian theory was resolved by R. A. Fisher in his [1918] using the multifactorial theory of inheritance.

[3] *Cf.* Burt and Howard [1956].

The data set out *below* in Table 1 are the correlations measured for pairs of variously related individuals. The data summarise correlations obtained on more than 30,000 pairs of individuals in over fifty separate studies carried out in eight countries. Most of the kinship correlations were obtained using very large groups although a notable exception is the rare category of monozygotic twins separated shortly after birth where a total of 122 pairs have been examined. The three kinships which have been investigated in a single study only[1] are second cousins (127 pairs), uncle or aunt and nephew or niece (161 pairs) and parents (as children) and their children (106 pairs). The results obtained experimentally are compared with those predicted on the assumption that the environment played no role at all in determining IQ's.

The theoretical values for the correlation coefficients for intelligence were calculated by Cyril Burt using the assortative mating parameter which he observed in his London sample. In applying these theoretical values to the correlations found for other samples, the assumption is made that there was, on average, a similar degree of assortative mating in these populations. This is a reasonable assumption since the theoretical values are rather insensitive to changes in the assortative mating parameter. (For example, increasing the value of this parameter from Burt's comparatively low value of 0.39 to 0.50, which is the value more usually found,[2] and keeping dominance the same, changes the theoretical correlation for pairs of siblings from 0.52 to 0.54.)

Some corresponding theoretical and observed correlations for stature are also set out in Table 1. The theoretical correlations for stature are estimated in the same way as are those for intelligence (that is, from the laws of genetics with the two parameters filled in).

The data presented in Table 1 show that there is a wide measure of agreement between the correlations found and those predicted for intelligence by the hereditarian programme and, in view of the large numbers of subjects involved, the agreement is highly significant. In fact the agreement between the theoretical and observed correlations is only slightly better for stature than for IQ.[3]

The correlations for IQ are compared with those for stature since the

[1] Burt [1966].

[2] For a summary of the correlations obtained for IQ between parents, *cf.* Jencks [1972], p. 272.

[3] There is also considerable agreement between observed correlations for IQ and those for finger-print ridges (a factor known to be almost entirely under genetic control). *Cf.* Huntley [1966].

latter character is one which has been used to test theories of inheritance and because stature is clearly not susceptible to those social-psychological and cultural influences which might be most plausibly thought to affect IQ scores.

TABLE I

CORRELATIONS FOR IQ BETWEEN RELATIVES

Relationship	*IQ*[1]		*Stature*[2]	
	Predicted	*Observed*	*Predicted*	*Observed*
Unrelated persons				
Children reared apart	0.00	−0.01		
Foster parent and child	0.00	0.20		
Children reared together	0.00	0.24	0.00	−0.07
Collaterals				
Second cousins	0.14	0.16		
First cousins	0.18	0.26	0.18	0.24
Uncle (or aunt) and nephew (or niece)	0.31	0.34	0.30	0.29
Siblings reared together	0.52	0.55	0.54	0.54
Siblings reared apart	0.52	0.47		
Dizygotic twins (different sex)	0.52	0.49		
Dizygotic twins (same sex)	0.52	0.56		
Monozygotic twins, reared apart	1.00	0.75	1.00	0.94
Monozygotic twins, reared together	1.00	0.87	1.00	0.96
Direct line				
Grandparent and grandchild	0.31	0.27	0.30	0.32
Parent (as adult) and child	0.49	0.50	0.51	0.51
Parent (as child) and child	0.49	0.56		

[1] The values for the observed IQ's are taken from Jensen [1969], p. 49; the theoretical values are calculated in Burt and Howard [1956] and Burt [1971].
[2] Burt and Howard, *ibid.*, p. 117.

Comparisons of physical characters, such as stature, with IQ scores play some heuristic role in the hereditarian programme. If IQ's behave in some unexpected manner and if the same phenomenon is observed in analogous results for some physical character, then the hereditarian is directed to try to solve the anomaly, not by changing the intelligence test but by looking for some genetic or biological explanation. Such a move led to a successful explanation of small deviations from normality in the distribution of IQ's.[1] In fact, in the earliest days of the hereditarian programme, before any satisfactory theory of inheritance was available, the comparison of mental and physical characters played a much more important heuristic role, since physical characters such as stature provided the standard of an inherited character. Thus Galton, in his earliest exposition of the hereditarian position, stated that 'the general resemblances in mental qualities between parents and offspring, in man and in brute, are every whit as near as the resemblance of their physical features; *and I must leave the existence of actual laws in the former case to the matter of inference from the analogy of the latter*'.[1] Thus, using stature as an analogy, Galton predicted that native ability is distributed approximately normally in the population[2] and he confirmed that 'artistic ability' is inherited by showing that parents resemble their children in this quality and in stature to similar extents.[3]

---

[1] *Cf. below*, pp. 128–30.

By considering the small discrepancies between the observed and theoretical correlations for IQ scores, it is possible to estimate to what extent the IQ tests measured inherited, genetic, differences. It is found, in fact, that for the IQ tests which were used, the amount of variance in scores which is attributable to environmental and genetic factors is about 20 and 80 per cent, respectively.[4] (Included in possible environmental factors are those which operate both pre- and post-natally and also random errors of measurement.) The above estimate for the importance of heredity in determining IQ differences is based on the assumption that genetic and environmental factors make independent, additive contributions to the total variance. Jinks and Fulker, in their reanalysis of Burt's (and others') data on children reared together and apart have shown that these assumptions are independently corroborated.[5]

It is immediately obvious on examining the correlation data in Table 1 that, from the environmentalist point of view, the IQ tests employed in the investigations did not tap any *inherited* intellectual capacity since they indicate that the more closely people are related the more alike they are in their IQ scores. (This result is of course predicted by the hereditarian programme.) However, since people who are genetically related are also brought up in similar surroundings, the relative sizes of many of these kinship correlations can be explained easily within the environmentalist programme as reflections of general social and cultural differences. For example, the relative sizes of the correlations found for siblings (reared together), cousins, and unrelated children are successfully accounted for by the fact that when siblings are brought up in the same home they generally share more socio-cultural factors than do cousins, and the latter usually experience more similar surroundings than do randomly selected people. For similar reasons, the resemblance between parents and their children in IQ score will usually exceed that between grandparents and their grandchildren.

---

[1] Galton [1865], p. 158; my italics.

[2] Galton [1869], pp. 31–2.

[3] Galton [1889], p. 218.

[4] *Cf.* Burt and Howard [1956], Jensen [1969] and especially Jinks and Fulker [1970]. Eysenck ([1971*a*], p. 63) notes that most of the correlation studies were undertaken using IQ tests which are less 'culture fair' than those which have been developed in recent years. He conjectures that genetic variance in individual scores on these modern tests would be higher than 80 per cent.

[5] Jinks and Fulker, *op. cit.*

Although the environmentalist programme can explain successfully *some* of the correlation data, taken as a whole these data constitute much greater empirical progress for the rival hereditarian programme. First, while the environmentalist theories only account for the *relative values* of some of the observed correlations, the hereditarian programme predicts precise quantitative values for each of the correlations and, moreover, as we have seen, the predicted values are very close to the observed ones. Secondly, among the observed kinship correlations, there are several which refute the environmentalist predictions while they confirm those of the hereditarian programme. The most striking of these refutations involve comparisons of (1) monozygotic twins brought up in different homes and unrelated children brought up together, and (2) monozygotic and dizygotic twins. I shall consider each of these refutations of the environmentalist theories in turn.

TABLE 2

Relationship	*Correlations for IQ*		
	Burt [1966]	*Newman* et al. [1937]	*Median value for several studies*[1]
Monozygotic twins (together)	0.92 (N = 95)	0.88 (N = 50)	0.87 (N > 1,000)
Monozygotic twins (apart)	0.87 (N = 53)	0.77 (N = 19)	0.75 (N = 122)
Unrelated children (together)	0.27 (N = 136)	—	0.24 (N = 251)

[1] These values are taken from Jensen [1969], p. 49.

(1) Since monozygotic twins share identical sets of genes, the hereditarian programme predicts that they possess identical degrees of inborn intelligence and hence have a pair-wise correlation for intelligence of 1.00, whether brought up together or not. On the other hand, the environmentalist programme predicts that if the twins are allocated randomly with respect to the environmental factors invoked to explain similarities between relatives, then these twins should be no more alike in their IQ scores than are unrelated children drawn at random. Table 2 contains a summary of the data obtained experimentally.

The data obtained by Burt, Newman *et al.* show that the correlations in IQ for monozygotic twins reared together are very high. In fact, the correlations are almost as high as those for IQ's measured at weekly intervals of the same individuals. Also, most of the studies give correlations for monozygotic twins who were separated early in life which are only slightly lower than those for monozygotic twins brought up together. Moreover, in contrast with the environmentalist prediction the median correlation of 0.75 for separated monozygotic twins is much higher than the median correlation of 0.24 for unrelated children brought up together. Although these data apparently corroborate the hereditarian theory, the 1937 study has often been held to corroborate the environmentalist theory.[1] Environmentalist claims in this respect have rested not on the observed overall correlation, but on certain specific aspects of the results; for instance, on the fact that among the 19 twins reared apart, 3 had IQ differences of at least 17 points. The largest difference was observed for 'Helen' and 'Gladys', one of whom was a teacher with an IQ of 116 while the other, who had lived in an isolated mountain district for most of her life, had an IQ of only 92.

Such differences constitute anomalies for the hereditarian programme. Burt suggested that since the twins had markedly different educations, the predominantly verbal tests were unsuitable.[2] While this explanation could not be tested on 'Helen' and 'Gladys' themselves, it gains some support from the fact that Burt obtained significantly higher correlations between the separated twins he studied when he used non-verbal tests of intelligence.[3] As striking as the case of 'Helen' and 'Gladys' is Burt's case of 'George' and 'Llewellyn'. While 'George' took a first class degree in modern languages, 'Llewellyn' was brought up on an isolated farm in North Wales and had reading and verbal abilities typical of a child of 11. Nevertheless, the twins scored 136 and 137 respectively on the non-verbal test.[4]

Both the high correlation for IQ of monozygotic twins reared apart and the low correlation which is found for unrelated children brought up together, constitute anomalies for the environmentalist programme which have not yet been dealt with, except in an *ad hoc* way. In order to account for the first anomaly, environmentalists have assumed that the environments of the separated twins which were studied were more similar than had been assumed. Similarly, the second anomaly has been explained by positing that the environments of the unrelated children were in fact different despite their superficial similarity.[5] These assumptions can only be made testable by specifying the precise (hidden) factors which made the environments alike in the first case and unalike in the second. But this has not so far been done and the explanations are consequently *ad hoc*. The environmentalist's task here seems all the more difficult in view of the fact that among the twins which Burt studied there was no correlation at all between the occupational statuses of the families to which each of the separated twins had been allocated.[1]

---

[1] *Cf. e.g.* Halsey [1958] and Maddox [1957].

[2] *Cf.* Burt and Howard [1956], p. 123.

[3] *Cf.* Conway [1958], p. 183, and Burt [1966].

[4] *Cf.* Conway, *ibid.*, p. 186.

[5] Lewontin ([1970], p. 6) advocates both of these *ad hoc* auxiliary assumptions.

(2) One of the most striking features of the family correlations for IQ is the fact that monozygotic and dizygotic twins are so different despite the similarity of the environments which these twins experience. The difference is *anticipated* by the hereditarian programme, since dizygotic twins, like siblings, have only 50 per cent of their genes in common, on average, while monozygotic twins have identical sets of genes. In order to account for this difference within the environmentalist programme, some workers have suggested that the environments of these twins differ in several crucial aspects. For example, monozygotic twins generally spend more time in each other's company, they are emotionally closer, and they are more likely to share their friends than are dizygotic twins.[2] However, the assumption that these factors influence IQ scores significantly is refuted by a comparison of the correlations for monozygotic twins reared apart and dizygotic twins reared together. Burt found a pair-wise correlation for IQ of separated monozygotic twins of 0.87, while the values predicted for this correlation by unqualified hereditarian and environmentalist theories are 1.00 and 0.00, respectively. Now for dizygotic twins of the same sex who were reared together, the correlation observed by Burt was 0.56, which is close to the value of 0.52 predicted by the hereditarian programme. The difference in the correlation coefficients for IQ of separated monozygotic twins and dizygotic twins reared together is over four times its standard error. In order to account for this difference, the environmentalist would have to assume that the separated monozygotic twins were subject to more similar environments than the dizygotic twins reared in the same homes. But no independent evidence has ever been brought forward to support this counter-intuitive assumption.

In predicting family similarities in IQ, the hereditarian and environmentalist programmes differ in two respects. First, the hereditarian programme makes precise quantitative predictions while its rival predicts only the relative values of *some* correlations between relatives and, secondly, the two programmes make several conflicting predictions. As I have shown, each of the hereditarians' quantitative predictions is close to the values obtained experimentally and, where the two programmes make opposing predictions, in particular those regarding unrelated children brought up together and related children reared apart, the hereditarians have been dramatically right while the refuted environmentalist theories were only rescued at the cost of their empirical content by *ad hoc* assumptions.

---

[1] Burt [1966].

[2] For a sympathetic review of some IQ-determining environmental factors which allegedly distinguish monozygotic and dizygotic twins, *cf.* Anastasi [1958], pp. 287–9.

(*b*) *Intelligence and Social Class.*

(*bi*) *Filial Regression to the Mean and Social Mobility.*

The hereditarian programme predicts that the correlation for intelligence between parent and offspring will be about 0.5. (As I have already discussed, the exact value for this correlation coefficient depends on the degree of assortative mating and of dominance exhibited by the population, and these are independently measurable.[1]) Now if, as is predicted by the hereditarian programme, the variability of IQ remains constant from generation to generation, then it follows from statistical theory that if there is a parent-offspring correlation of 0.50, then children have, on average, an IQ lying mid-way between the parental IQ level and the level of the general population. Thus, fathers or mothers with average IQ's of 120 will have children whose mean IQ level is only about 110 and those with average IQ's of 80 will have children whose mean IQ is about 90. This phenomenon is called filial regression to the mean. It is typical of characters generally assumed to be largely genetically determined, such as finger-print ridges and stature.

Now consider a population divided into hypothetical social groups which are differentiated by the average IQ of their male adult members. According to the hereditarian programme, (*i*) the intelligence of the children within each social group will regress half-way towards the mean of the general population and (*ii*) the standard deviation of the children's intelligences within each group will be close to that of the general population. The hereditarian programme predicts therefore that if the frequency distribution of IQ's within each of the social groups remains constant from generation to generation, some of the children will move out of the social groups to which their fathers belong. The minimum social mobility which is required to maintain the social distribution of IQ's can be easily calculated.[2]

Now socio-economic classes are similar to the hypothetical groups just considered. It has been found universally since the earliest days of mental testing that there are significant differences in mean IQ between the various socio-economic classes.[3] Burt's studies, for example, show that when occupational classes are arranged in the order of 'the apparent difficulty of the mental processes required either to carry out the occupation efficiently or to gain entrance into the occupation',[1] then the average IQ ranged from about 140 for the higher professional group to about 85 for unskilled workers.[2]

---

[1] *Cf. above*, p. 115.

[2] *Cf.* Burt [1961*a*].

[3] *Cf. e.g.* Binet [1911], pp. 187–8.

On the assumptions (*i*) that the IQ tests are perfect measuring instruments for inborn differences in intelligence, (*ii*) that the frequency distribution of IQ's within each class remains constant from generation to generation, and (*iii*) that a person's intelligence is the sole factor determining whether he will be socially mobile, Burt predicted that on average 22 per cent of children will move out of their father's occupational class.[3] The first of these assumptions is independently confirmed by the correlation data for variously related people already discussed[4] and the second is corroborated by data collected by Burt showing that over the period 1920–50 with which he was concerned the 'occupational distribution of intelligence has remained fairly constant'.[5] The third assumption is not independently supported and consequently Burt's prediction must be weakened to the assertion that the average level of social mobility is *at least* 22 per cent.[6] Now the observed amount of social mobility in England during the period considered by Burt was about 30 per cent,[7] so that in its weakened version, the hereditarian prediction is confirmed.

Now the approximate level of social mobility in England was already known when Burt deduced its minimum extent from hereditarian theory. The prediction was therefore not novel in the temporal sense. However, it was an entirely unexpected outcome of the hereditarian programme; that is, it played no heuristic role in the construction of the hereditarian theory from which it was deduced. After all, Burt's prediction would not have been any different if he had been entirely ignorant of the empirically ascertained level of social mobility. The prediction is therefore novel in the sense defined recently by Zahar.[8]

But the prediction is not important in the context of the 'IQ debate' *simply* because the observed level of mobility does indeed exceed the minimum required, for the prediction that there is at least 1 per cent mobility would have also been novel in Zahar's sense, although intuitively it would have been much less impressive. The dramatic success of the prediction of 22 per cent mobility lies in the fact that it is not only confirmed but that no non-hereditarian theory could have independently anticipated so much social mobility. This may be judged from the *ad hoc* conjectures which have been put forward by those who eschew genetic theories when explaining social mobility.

---

[1] Burt [1959], p. 19.
[2] Burt [1961*a*] and [1943].
[3] Burt [1961*a*].
[4] *Cf. above*, section 2(*a*).
[5] Burt *ibid.*, p. 15.
[6] The level of social mobility is the proportion of people who have moved out of the class to which their fathers belong. For the purpose of his calculations, Burt divided the male population into three separate classes. Class 1 includes professional, administrative and clerical workers; class 2 includes skilled and semi-skilled workers; and class 3 includes unskilled workers.
[7] *Cf.* Burt *ibid.* and Glass [1954], p. 183.
[8] Zahar [1973], pp. 101–4.

According to one influential writer on social mobility, upward mobility 'is the product of *fluidity* and opportunity, changes in the occupational distribution, as well as demographic changes, [while] downward mobility is also a product of these factors, but here *social fluidity* is the more important'.[1] This hypothesis is, however, robbed of most of its empirical content in a footnote where the term 'fluidity' is defined as 'the ease of movement from stratum to stratum in society that is not due to changes in the occupational structure'.[2] In explaining the present level of social mobility, two of the leading authorities on the subject refer to the differential fertility amongst the social classes, the changing ratio of white collar and manual jobs and to 'motivational' factors.[3] But none of these factors accounts for any important amount of mobility. Taking Glass's empirical data for social mobility in Great Britain, the differential fertility of manual and non-manual workers caused less than 5 per cent mobility,[4] and the changes in the occupational structure made no significant contribution to mobility.[5] Since the nature of the 'motivational' factors is not spelled out, their importance for mobility cannot be estimated.

It is clear then that while hereditarian theories account for most of the observed mobility in a progressive way, there is at present no satisfactory environmentalist answer to the question about the causes of social mobility.

While the successful prediction of social mobility is a dramatic novel fact for the hereditarian programme, the phenomenon of filial regression to the mean on which this prediction depends is an anomaly for the rival programme. For if class-correlated environmental factors were crucial in creating IQ differences, environmentalists could not explain why so many children of unskilled workers gain IQ's above the population average and so many of those born to 'professional' parents are below the average in IQ. In order to explain such phenomena within the environmentalist programme, it would be necessary to invoke some factors which are *negatively* correlated with social class, but so far no environmentalist has used such factors to account for the phenomenon of regression to the mean. However, Eysenck has suggested the possible explanation that upper class children, while they have all the advantages that money can buy, are deprived of those non-material factors such as intellectual companionship and parental concern for the child's well-being which are also important in determining IQ. Lower class children, on the other hand, while they suffer the disadvantages of relative poverty, are spared the burden of parents who spend all their time at cocktail parties and European conferences and hence they enjoy those non-material IQ determinants which are lacking in upper-class homes.[1] Eysenck has expressed the view that it would be 'illogical' for environmentalists to claim both that the higher classes have higher IQ's because of their advantageous environment and to claim that regression in these groups arises from disadvantageous factors.[2] But, *pace* Eysenck, these explanations for the class distribution of IQ's and for regression are perfectly consistent so long as the *same* environmental factors are not deployed to explain *both* facts. The environmentalist may claim that children have a 'natural' IQ of about 100 and that this value is changed by certain environmental factors. In the case of middle class children, the 'natural' IQ is increased by virtue of their superior education and material environment but this increase falls short of reaching the parental level because of certain non-material factors. *The weakness of this explanation for regression to the mean is not that it is inconsistent but that it is* ad hoc. The environmentalist must assert that the material and non-material factors just happen to be balanced in such a way that on average a child's IQ regresses almost exactly half-way towards the population mean. This arbitrary adjustment of the environmental parameters would clearly be *ad hoc*.

---

[1] Miller [1960], p. 59.

[2] *Ibid.*, p. 59.

[3] Lipset and Zetterberg [1956].

[4] Glass [1954], p. 197.

[5] *Ibid.*, pp. 189–94. In contrast with several other industrial societies, the levels of upward and downward mobility in Britain are 'virtually the same'. *Cf.* Bendix and Lipset [1960], p. 289.

According to the hereditarian programme, the differences in average IQ between the classes are due to genetic factors. Hereditarians assume that these differences arose through the accumulated effects of social mobility—that is, by the selective migration of the more intelligent to a higher occupational class.[3] Halsey, who advocates 'the hypothesis of near-randomness in the present social distribution of innate intelligence',[4] maintains that this theory of selective class migration is unreasonable since it predicts that there has been a significant amount of social mobility since as early as 1700.[5] Halsey, however, offered no evidence that social mobility is a novel phenomenon; but the prediction that there was extensive social mobility in the past has not been confirmed either.[6]

Nevertheless, if the social class differences in IQ are the result of the accumulated effects of social mobility, the hereditarian programme predicts that they are maintained by this process. This prediction is confirmed by the result discussed above which shows that the true level of social mobility (30 per cent) is in excess of that required to maintain the present social class distribution of IQ's (22 per cent).

---

[1] Eysenck [1971*a*], pp. 68–9.

[2] *Ibid.*, p. 69.

[3] *Cf. e.g.* Conway [1959].

[4] Halsey [1959], p. 1.

[5] Halsey [1958] and [1959].

[6] Burt does however cite some anecdotal evidence taken mainly from Galton's [1869] investigations showing that many prominent historical figures rose to high social positions from poor origins. *Cf.* Burt [1959], p. 24, and Burt [1961*b*], pp. 124–8.

The correlation between IQ and occupational class is significantly below unity and many of those in even the lowest occupational group gain IQ scores which are well above the average for the population. Herrnstein has predicted that if contemporary meritocratic ideals are fully realised and occupational class becomes more closely linked with innate ability, then the IQ differences between the classes will widen and the level of social mobility will eventually fall. One of the predicted consequences of this is that social classes will take on the character of castes.[1] This rather alarming prediction has not been corroborated.

(*bii*) *Dysgenic Effects.*

Since it had been discovered that families of low social class are on average larger than families of high social class and since it was assumed that the positive correlation between social class and IQ is genetic in origin, hereditarians have predicted that there will be a gradual lowering of the mean population IQ by about 1 point per decade.[2] However, when these predictions were tested using non-verbal tests of intelligence in a large-scale investigation of ten-year-old children in Leicester, the anticipated decline was not found. In fact, over thirteen years from 1936 to 1947 there was a small but significant increase of 1.28 points in the population's mean IQ.[3] In a similar study of the Scottish population, where verbal tests were employed, there was a rise in the population IQ of 2.28 points over the period 1932–47.[4]

Hereditarians have attempted to explain these anomalous results by three auxiliary assumptions. First, the anomaly has been ascribed to inadequacies in the tests used and Cattell suggests that there was a larger gain in mean IQ in Scotland because of the use of relatively culture biased verbal tests.[5] Secondly, it is said that the fall in intelligence was masked by an 'advance in education and test sophistication sufficient to produce an upward change in test score of about twice the magnitude of the downward genetic change'.[1] And thirdly it was suggested that 'the differential birth-rate (favouring those of low IQ) has been offset by differential death, celibacy, barren marriage, etc.'[2] These auxiliary hypotheses are all *ad hoc*. But there is some evidence which supports the assumption that, although the family size of the lower social groups is, as a rule, relatively high, this effect is sometimes compensated by the higher frequency of celibacy amongst those of low IQ.[3] But since these results were obtained from populations quite different from those in which the anomalous IQ increases were observed many hereditarians still regard the anomaly as unresolved.[4]

---

[1] Herrnstein [1971] and [1973].

[2] Cattell [1937], pp. 42–3. A decline in intelligence caused by differential rates of breeding has been predicted by hereditarians since the end of the nineteenth century and the fear of such a decline was a factor in the founding of the Eugenics Society in 1908 under Galton's presidency. In this century, many hereditarians have warned against the dire consequences of the observed differential birth-rate. For example, Cattell (*ibid.*, p. 43) predicted that if the effects of a differential birth rate are not countered 'half the population would be mentally defective' in three hundred years.

[3] *Cf.* Cattell [1950].

[4] Thompson *et al.* [1949].

[5] Cattell, *ibid.*, p. 141.

The methodology of scientific research programmes is distinguished from some other philosophies of science such as naive falsificationism and inductivism in that it allows that a programme may be progressively scientific and yet incorporate unresolved anomalies. Thus, in appraising a programme using the standards of this methodology, the crucial task is not to see whether it has run up against any anomalies—nearly all programmes are anomaly-ridden—but rather to see whether it is dramatically confirmed, that is whether it has predicted any novel facts. Galton appreciated the insignificance of anomalies and the paramountcy of dramatic confirmation when he first set out the central tenets and outlined the heuristic of the hereditarian programme. He wrote: 'Resemblance [between parents and their children in mental qualities] frequently fails where we might have expected it to hold; but we may fairly ascribe the failure to the influence of conditions that we do not yet comprehend. *So long as we have a plenitude of evidence in favour of the hypothesis of the hereditary descent of talent, we need not be disconcerted when negative evidence is brought against us.*'[5]

The small increase in the population's mean IQ over 13 years is rather easily explained in environmental terms since there has been a general improvement in prosperity in the areas concerned over this period. However, since the rise in IQ was only about 1 point while the difference in mean IQ between children of the highest and lowest classes is around 27 points,[6] this result constitutes an extremely modest success for environmentalist explanations of social class differences in IQ.

The hereditarian programme does not predict the existence of any differences between the classes in inherited intelligence. But since the assumption that IQ tests largely measure inborn differences is corroborated in the studies of correlations for IQ of relatives reared together and apart,[7] hereditarians assume that the observed class distribution of IQ's reflect, in the main, inherited, genetic differences. On the basis of this assumption, and using other auxiliary assumptions, the hereditarian programme makes several predictions: (*i*) there will be a certain minimum amount of social mobility,[1] (*ii*) there was considerable social mobility in the past,[2] (*iii*) the social classes will gradually take on the character of castes,[3] and (*iv*) there will be a steady decline in the population's mean IQ.[4]

---

[1] *Ibid.*, p. 140.

[2] *Ibid.*, p. 140.

[3] *Cf.* Bajema [1963] and Higgins, Reed and Reed [1962]. Also, *cf.* Carter [1966] for a discussion of evidence relating to the effects of differential fertility.

[4] *Cf. e.g.* Jensen [1969], p. 94.

[5] Galton [1865], p. 158; my italics.

[6] *Cf. e.g.* Burt [1961*a*].

[7] *Cf. above*, section 2(*a*).

As I have shown, the first prediction was dramatically corroborated, the second and third have not yet been tested and the fourth was disconfirmed. Although environmentalist theories can successfully account for the observed small rise in the average population IQ, they cannot explain the phenomenon of social mobility, nor that of filial regression to the mean. Of course, environmentalists reject the hereditarian assumption that there are any inherited intellectual differences between the classes and I shall discuss the *ad hoc* environmentalist explanations for the social class differences in IQ in section 3.[5]

(*c*) *The Distribution of Intelligence.*

The hereditarian programme predicts that intelligence is ***normally distributed.*** This prediction is a consequence of the assumption that intelligence is inherited by the multifactorial mechanism discussed above.[6] Since the segregation and recombination of genes is a matter of chance, the frequency distribution of the character they influence will be normal. In fact, most graded physical characters, for example stature and lung capacity, are distributed approximately normally. (As I mentioned above, Galton predicted the Gaussian distribution of native ability on the basis of an analogy with stature.[7])

The hereditarian prediction is confirmed by the fact that IQ distributions conform closely to a normal curve within the IQ range 70 to 130, a range which encompasses about 95 per cent of the population. There are, however, a significant departures from normality outside the range 70 to 130. For example, the number of children with an IQ above 160 exceeds the number predicted by a factor of more than ten[8] and the percentage of children whose IQ's lie between 20 and 50 is four times the anticipated percentage.[9]

Hereditarians have successfully explained the fact that IQ's are not quite normally distributed by the auxiliary assumption that while the intelligence of most people is determined by the combined action of a large number of genes each of which has a small effect, some extreme levels of intelligence are the product of rare, single, recessive or mutant genes which affect intelligence in a major way.[1]

---

[1] *Cf. above*, p. 123.

[2] *Cf. above*, p. 125.

[3] *Cf. above*, p. 126.

[4] *Cf. above*, p. 126.

[5] *Cf. below*, Part II.

[6] *Cf. above*, p. 115.

[7] *Cf. above*, p. 118.

[8] *Cf.* Burt [1963], p. 182.

[9] *Cf.* Penrose [1963], p. 45.

The origin of this hypothesis illustrates the heuristic role which comparisons of intelligence and stature play within the hereditarian programme.[2] The hypothesis was suggested by the fact that similar deviations from a Gaussian distribution curve were observed in the case of stature. As early as 1917, Burt argues that the elongated tail of the normal curve is due to 'the existence of a distinct species of pathological "defectives" '[3] and he says that 'for confirmation [of the theory that the observed distribution reflects the true distribution of inherited intelligence] it would be most suggestive to compare figures for physical variability'.[4]

The assumption that very low IQ's are usually caused by single, mutant or recessive genes was dramatically confirmed in a study of 271 'mentally defective' (IQ 30–68) children and their 562 siblings.[5] The hereditarian programme makes the following predictions concerning the 'mentally defective' subjects and their siblings: (*a*) the subjects will be either members of a *low* IQ 'imbecile' group or of a *higher* IQ 'feeble-minded' group, depending on whether their level of intelligence was primarily the result of single genes or whether it was determined multifactorially (as it is for the general population). (*b*) Since the intelligence of the 'imbecile' subjects is determined mainly by rare, single genes, their siblings will not usually share the same gene. Consequently, the siblings of the 'imbeciles' will have IQ's which are representative of the general population. That is, their IQ's will be normally distributed and have an average value near 100. (*c*) On the other hand, since the intelligence of the 'feeble-minded' subjects is determined in a multifactorial mechanism, they will on average have 50 per cent of their genes in common with their siblings. As a result, the siblings will have a mean IQ about mid-way between the mean population IQ and the average IQ of the 'feeble-minded' subjects.[6] (*d*) Since socio-economic status is correlated with intelligence in the general population[7] and since the hereditarian programme predicts that the intelligences of the 'feeble-minded' and of the general population are both determined by the same mechanism, the 'feeble-minded' subjects will be mainly from the lower social classes.

Each of these predictions was confirmed in the study of 'mentally defective' children. Although the IQ's of the siblings as a whole were not normally distributed, an 'imbecile' group (mean IQ about 49) had siblings whose IQ's were approximately normally distributed with a mean value close to 100, while a 'feeble-minded' group (mean IQ about 58) had siblings whose IQ's were normally distributed about a mean value of about 80.[1] Moreover, the 'imbecile' subjects came from an average cross-section of social backgrounds, while the 'feeble-minded' subjects were predominantly from the lower social classes.[2]

---

[1] *Cf.* Burt [1935], pp. 79–81.

[2] *Cf. above*, pp. 117–8.

[3] Burt [1917], p. 35.

[4] *Ibid.*, p. 31.

[5] These 'mentally defective' children were, supposedly, all those with a Stanford–Binet IQ in the 'mentally defective' range in the Bristol and Colchester areas in England. Roberts [1952].

[6] For the phenomenon of filial regression to the mean, *cf. above*, pp. 122–5.

[7] *Cf. above*. p. 122.

The fact that intelligence is distributed in an approximately normal fashion represents a dramatic victory of the hereditarian programme over rival theories. First, it contradicts the supposition which, according to Thorndike, was commonly held at the beginning of this century that the groups designated 'genius', 'normal', 'feeble-minded', 'imbecile' and 'idiot' form completely separate classes and that the distribution of ability should therefore be ***multimodal***.[3] Secondly, no environmentalist theory has been able to provide anything but an *ad hoc* explanation for the normal distribution of IQ's. In order to explain this distribution, environmentalists would have to assume that there are a large number of independent environmental factors each of which makes a small contribution to an individual's final IQ score. But this assumption contradicts other theories which have been put forward by environmentalists to explain social class and racial differences.[4] In these cases, the environmental factors usually suggested as the most important are not independent, but are linked with socio-economic status. If these class-correlated factors were important, then IQ distribution curves should be skewed towards the lower end as are the curves for school attainments and incomes. According to one environmentalist, the assumption that the relevant environmental factors are independent is ruled out by sociological considerations since 'views of reality and values are structured [and] they hang together amongst groups of people rather than being randomly distributed'.[1]

---

[1] According to Jensen, the subjects were divided into 'imbecile' and 'feeble-minded' groups according to whether their IQ's fell below 50 or in the range 50–75 ([1969], p. 26). But in fact there was no independent means for identifying the members of the two groups and a number of those classified as 'imbecile' had IQ's above 60. Whether the subjects had been successfully separated into those whose IQ's were part of the normal distribution and those whose IQ's were mainly the effects of major, single genes depended on whether predictions (*b*), (*c*) and (*d*) were confirmed. Roberts gives the number of 'imbecile' and 'feeble-minded' subjects in the IQ ranges 30–53, 53–60, 60–8. In estimating the mean IQ's of the subjects in the two groups, I have assumed that those in the three IQ ranges have mean IQ's of 41.5, 56.5 and 64 respectively.

[2] Hereditarians have achieved no comparable success in providing evidence for a distinction at the higher end of the scale between people with normal and non-normal IQ's.

[3] *Cf.* Thorndike [1914], pp. 315–46. Terman also claims that the approximately normal distribution of IQ's is a victory for the hereditarian programme since it refutes 'the *common opinion* that extreme deviations below the median are more frequent than extreme deviations above the median' (Terman [1916], p. 67; my italics).

[4] Attempts to explain race differences in average IQ in environmental terms are discussed in Section 3 (Part II).

Some environmentalists suggest that the normal distributions of IQ's cannot be counted a novel fact for the hereditarian programme since (they allege) it is not an empirical fact at all but merely a 'statistical artifact'.[2] According to this view, 'the normal distribution is a function of the particular mathematical methodology employed *and nothing else*'.[3] Another objection is that the normal distribution was (allegedly) secured by merely 'tinkering with the test material'[4] and that this (alleged) *ad hoc* rejection of test items if they did not yield the desired distribution made the normal distribution a '*foregone conclusion*'.[5] If one of these criticisms is justified, then the other is clearly redundant since if the normal distribution could be rigged by mathematical manipulation *alone*, there seems little point in 'tinkering with the test material'. Moreover, the two objections are separately unjustified. First, it is untrue that *any* distribution of IQ scores can be converted into normal form unless the number of people passing the various tests is forged. Also, if mathematical procedures *by themselves* could ensure a normal distribution, then the small deviations from normality would presumably have to be put down to a mathematical ineptitude on the part of the test constructors. Secondly, the objection that the (alleged) *ad hoc* elimination of test items made the normal distribution a '*foregone conclusion*' is also wrong, since even if researchers *had* set out to obtain a particular distribution of IQ's, there was no divine guarantee that their efforts would have been successful. As I have shown, researchers were not able to construct intelligence tests by using the heuristic rules of the faculty theory, even though they tried very hard.[6]

Nevertheless, it must be admitted that, if the prediction that intelligence is distributed normally had been confirmed by the *ad hoc* elimination of test material or after the arbitrary scaling of test scores, the environmentalist criticisms would be substantiated. For hereditarians could then only claim a victory after they had explained in a non-*ad hoc* way why the eliminated material was anomalous or why the original method of scaling had failed.

However, the prediction that intelligence is normally distributed was not confirmed in the *ad hoc* fashion alleged by environmentalist critics. Consider Binet-type tests. In these, various tests are allocated to successive 'age levels' and the child is credited with a mental age between the 'age level' at which he passes all the tests and the 'age level' at which he fails them all. Test items passed at intervening 'age levels' contribute fractions of a year to his mental age. Now IQ tests are standardised in a trial-and-error fashion by placing test items so that for each age group the average mental and chronological ages are identical and, hence, so that the population has a mean IQ of 100. In calculating mental ages for the purpose of this standardisation, children's performances are compared with the performance of children at *successive* age levels and no assumption is made about the relative performance of children *within* a particular age group.[1] Nevertheless, it turns out that within each age group, the distribution of Binet IQ's lies close to a normal curve, as predicted by hereditarian theory.[2]

---

[1] Swift [1972], p. 155.

[2] Lewis [1957].

[3] Daniels and Houghton [1972], p. 74; my italics.

[4] Richmond [1953], p. 228.

[5] Simon [1971], p. 67, footnote; my italics.

[6] *Cf. above*, pp. 112–3.

We see then that the hereditarian programme has not only successfully predicted the fact that intelligence is normally distributed, but it has also provided a content-increasing way out of a difficulty posed by small deviations from normality. In contrast, environmentalists have either denied the facts or dealt with them in an *ad hoc* fashion.

(*d*) *Racial Differences in Intelligence.*

Significant differences in the mean IQ scores obtained by members of different racial groups were first discovered in the course of an extensive investigation of American army recruits during the First World War. The particular racial difference which has attracted most attention from researchers is the approximately 15 point discrepancy in the average IQ scores of black and white Americans.[3]

According to the environmentalist programme, all racial groups have the *same* inherited abilities and hence any IQ differences must arise through some environmental factors. Attempts to provide independent evidence for this assumption will be discussed in section 3.

Hereditarians, on the other hand, consider that the observed difference in mean IQ score of American blacks and whites is largely a product of the different genetic endowments of these groups.[4]

The hereditarian programme does not predict whether any *particular pair* of racial groups will differ in their mean intelligence levels. However, many hereditarians argue that since races commonly differ in other polygenically determined physical characters, such as stature and skin pigmentation, it is plausible to assume that there will be *some* racial differences in intelligence.[1] Indeed, from the hereditarian point of view, it is reasonable to assume that some races will differ in their mean intelligence levels. For even if all groups had been endowed *ab initio* with similar levels of intelligence, there are so many social forces, such as selective migration and selective rates of breeding and mortality, which could alter the original racial distribution of intelligence that if all groups had retained the same mean intelligence levels, hereditarians would have to make the counter-intuitive assumption that either each group had experienced these social forces in the same degree or that such forces had operated so that their overall effects had cancelled out...

---

[1] For a discussion of the methods used to calculate mental ages, *cf. e.g.* Freeman [1963], p. 204.

[2] Those involved in constructing the Stanford–Binet intelligence tests state explicitly that 'no attempt was made, as erroneously claimed by some, to secure a normal distribution of the resulting IQ's. Items were chosen in such a manner that the average *M.A.* [mental age] for an age group coincided with their (*sic*) *C.A.* [chronological age]'. (McNemar [1942], p. 17.)

[3] For the most comprehensive review of the relevant studies, *cf.* Shuey [1966].

[4] Some people think that the research into the intelligence of whites and blacks constitutes two entirely separate research programmes. I shall discuss this erroneous view in section 3 (*b*) (Part II).

In this section, I shall discuss two predictions made by the hereditarian programme with respect to racial differences in intelligence.

The hereditarian programme predicts that if two racial populations differ with respect to their average levels of intelligence, then interbreeding between the two populations will produce offspring whose average intelligence lies between that of the two groups. Methods for measuring the degree of white ancestry of individual Negroes have only recently been worked out however and no attempts have yet been made to use these methods to correlate IQ's with the degree of racial admixture.[2]

The hereditarian programme predicts that intelligence is inherited in accordance with the multifactorial mechanism in all populations. It consequently predicts that within each population, provided that there is a similar degree of assortative mating and of dominance, the correlations for inherited intelligence for siblings are the same. In particular, assuming dominance and assortative mating to be close to the values generally found in white populations, the predicted sibling correlation for both blacks and whites is close to 0.50. Now since the hereditarian programme predicts that individuals and their siblings have identical, normal distributions of intelligence, it also predicts that children will have siblings whose intelligence has, on average, regressed half-way towards the population mean. More especially, it is predicted that the siblings of both black and white individuals will, on average, regress about half-way towards their own population means of about 100 and 85, respectively. This prediction has been corroborated.[1] When black and white children were matched for IQ, Jensen found that over the IQ range 50–150, the siblings of black children had on average an IQ 7–10 points lower than the siblings of white children. Thus, for example, black and white children matched for IQ 120 had siblings whose average IQ's were about 100 and 110 respectively. Similarly, Negro and white children with IQ's of 70 had siblings with mean IQ's of 78 and 85, respectively.

---

[1] *Cf. e.g.* Galton [1869], p. 351, Jensen [1973*a*], pp. 130–1, and Morant [1956], p. 320.

[2] Several investigations have shown that skin colour correlates positively with IQ in Negroes. But since skin colour is a rather poor predictor of the degree of mixed racial ancestry and since a person's skin colour has important social effects, these results do not constitute a satisfactory test for either of the rival IQ programmes. For a review of studies relating IQ and skin colour, *cf.* Shuey [1966], pp. 452–66 and Jensen [1973*a*], pp. 219–30.

So far, no non-*ad hoc* environmentalist account has been offered for this striking result. However, there is one striking *ad hoc* theory. According to Thoday, although the data on which the different sibling regressions are based are impeccable, they 'add nothing whatsoever to the strength of the genetic [that is, hereditarian] hypothesis'[2] since it is as easily susceptible to explanation by positing the existence of a certain '*environmental factor X*' which is specific to the black population and which penalises *all* members of that population *equally*.[3] Thus, if the effect of 'environmental factor *X*' would be eliminated, the IQ's of all the Negroes would be uniformly increased by about 15 points and the sibling regression data for blacks and whites would be brought into line. Although he tentatively suggests some candidates for the mysterious 'factor *X*', Thoday provides no independent evidence whatsoever that these are relevant to the issue of IQ differences nor does he indicate how such evidence might be obtained. But for Thoday the lack of independent empirical support for a conjecture is no deficiency since he considers that 'it is not reasonable to discount these [*X* factors] simply because they present hypotheses difficult to test, and because some other hypothesis fits the data'.[4] It is true that both an hereditarian and environmentalist hypothesis 'fit the data'. What distinguishes these hypotheses is that the first predicted the different regression effects for black and white siblings while the second was concocted *ad hoc* after the result was already known. Any data in the world can be made consistent with any theory by invoking nameless and untested factors. *Thoday's hypothesis is completely untestable and hence completely pseudoscientific.*

(*e*) *Conclusion.*

The hereditarian programme has anticipated many novel facts. I have discussed the successful predictions (*i*) of the degree of family resemblances in IQ[5], (*ii*) of IQ-related social mobility,[6] (*iii*) of the distribution of IQ's,[7] and (*iv*) of the differences in sibling regression for American Negroes and whites,[1] since these are the most striking examples. When the environmentalist programme has attempted to account for the novel facts produced by the hereditarian programme, it has been unable to do so except in an *ad hoc* fashion. The hereditarian programme is not free from anomalies—but in this respect it is not exceptionally placed among scientific research programmes. Anomalies are ***anticipated*** by a good research programme and the puzzle-solving techniques for resolving them are articulated in advance. The crucial criterion for deciding whether a programme is progressive is whether it has anticipated novel facts. By this criterion, the hereditarian programme has been scientifically progressive and has contributed to the growth of knowledge.[2]

---

[1] Jensen [1973*a*], pp. 117–19.

[2] Thoday [1973], p. 419.

[3] *Cf.* Jensen [1973*a*], pp. 137–9, for a discussion of environmental factors of this type and of their *ad hoc* nature.

[4] Thoday, *ibid.*, p. 419.

[5] *Cf. above*, pp. 115–22.

[6] *Cf. above*, pp. 122–5.

[7] *Cf. above*, pp. 128–30.

(*To be continued*)

The London School of Economics

---

[1] *Cf. above*, pp. 133–4.

[2] Some people consider that on Lakatos's criterion the hereditarian programme is pseudoscientific: I am unable to see the force of their argument. (*Cf.* Richardson and Houghton [1973], p. 160.)

Brit. J. Phil. Sci. 25 (1974), 137–154 Printed in Great Britain

Pattern Recognition in Computers and the Human Brain:

With Special Application to Chess Playing Machines

by ROLAND PUCCETTI

Introduction.
1 *Matching Templates and Feature Analysers.*
2 *Modes of Perception in Left and Right Cerebral Hemispheres.*
3 *Identification and Recognition.*
4 *Chess Playing Machines.*

INTRODUCTION

In the opening pages of his book on the subject of mechanical pattern recognition, Sayre ([1965]) sounds the following pessimistic note:

> We have no clear notion of what should be done to gain significant improvement in the performance of our simulation mechanisms. The reason for this is that we do not know exactly what is going wrong; and the reason for this in turn is that we lack appropriate criteria for judging the adequacy of our present results. Behind these methodological obscurities, however, is an unclarity about the nature of the human behavior we are trying to simulate. We simply do not understand what recognition is. And if we do not understand the behavior we are trying to simulate, we cannot reasonably hold high hopes of being successful in our attempts to simulate it (p. xii).

Gunderson ([1971]), while granting that pattern recognition remains perversely 'program-resistant', thinks it a mistake to become discouraged on such grounds.

> Sayre seems to have assumed that only after we have clarified the concept of recognition will we understand how to program a machine with recognition capacities, whereas I want to suggest that by trying to program a machine with recognition capacities we will perhaps improve our understanding of the concept of recognition—and, of course, of recognition itself (p. 92).

But of course if pattern recognition in humans is not achieved by means which lend themselves to machine simulation, then neither shall we succeed in simulating it mechanically nor shall we learn from our failures how we do it ourselves.[1] And this could have considerable practical importance. One of the reasons computer simulation of pattern recognition has been so vigorously pursued and generously funded is that it holds the

Received 8 November 1973

---

[1] If birds were intelligent enough they might have developed 'flying machines' like ours. But apart from revealing the law of aerodynamic lift over an airfoil, such machines would

promise, if successful, of developing automatic mail sorters, rapid machine translation, even mechanical readers for the blind. If such attainments are seen to be unlikely, not just because our technology is insufficiently advanced but because it is misguided, we shall at least save ourselves a lot of effort and money.

I MATCHING TEMPLATES AND FEATURE ANALYSERS

In classical psychology the two main theoretical approaches to pattern recognition have been 'template-matching', in which each new stimulus input is supposed to be compared to a stored standard; and 'feature-analysis', where the presence of particular parts or properties of the stimulus is thought to be decisive for recognition. Let us take 'template-matching' first. Neisser ([1967]) explains it as follows:

> A new figure is identified by noting its coincidence, or congruence, with a basic model. Template-matching is not uncommon in daily life. To determine if a particular fingerprint was made by one of his suspects, a detective may superimpose it, successively, on the prerecorded prints of each one until he obtains a match. Even then, the new print and the standard will probably not be quite identical: the sample will be smudged or distorted or partially incomplete. Nevertheless, the detective will be satisfied as long as the match is reasonably good, and better than any other. In a sense, he computes a correlation coefficient between the two patterns and asks if it is near 1.00 (p. 50).

There are obvious difficulties with this theory. For one thing, the careful comparison of details suggested in Neisser's analogy is only rarely characteristic of how we recognise a familiar object or face. I might use the detective's technique if I had to pick out, say, my car from a lot containing several of the same make and model, or a criminal in a police line-up; but this cannot be the way I recognise my wife in the morning or my desk when I arrive at the office. But there are other problems. As Höffding ([1891]) pointed out long ago, on this theory the slightest rotation, displacement, or even change of size in the stimulus input would result in a mismatch with the template, whereas in fact we have no difficulties recognising a form even when it appears in a new retinal position, despite size changes or rotations. In Hebb's ([1949]) words, 'it is hard to reconcile an unlocalised afferent process with a structural (and hence localised) mnemonic trace' (p. 15).

As might be expected, attempts to simulate human pattern recognition by machine on the template-matching model have been of very limited

---

not simulate the flight of birds, and could not teach their avian designers how they fly. The argument that this does not matter since such machines fly better than birds can be countered by the observation that for most purposes for which birds fly that is not true: even a helicopter has nothing like the mobility of birds.

success. Ways to overcome the difficulties just mentioned have been devised, namely by 'preprocessing' the stimulus before it is compared to a template. In brief, this consists of (*1*) 'cleaning up' the input so as to eliminate distracting imperfections in the photograph or printed material presented to the machine, such as holes or isolated points; and (*2*) 'normalising' it by centring it on the template, expanding or contracting it to a standard size, and rotating the figure until its longest axis reaches a fixed orientation. Such refinements have made computer recognition workable with a limited range of patterns, but the results are still a far cry from what humans can do. As Neisser ([1967]) points out, ***ill-defined*** patterns, those lacking clear-cut boundaries and hence unsuitable for matching against a template—such as the visual distinction between a cat and a dog—are the rule rather than the exception in our daily lives (p. 58). Even in something so simple as hand-printed letter recognition there could be no single template adequate to match all the forms humans easily recognise as an '*A*' without any use of context (Neisser and Weene [1960]). And finally, the template theory does not allow for the fact that tiny details often determine recognition of a form for what it is, no matter how much its general characteristics overlap with a different one. Neisser provides the following example:

> The difference between a *Q* and an *O* is minute compared with the shape differences between *Os*, but it is decisive. The curvature of a small line segment can alter the expression of a cartoon face completely. The theoretical significance of this observation is not that the category is ill-defined (though it may be, in the case of facial expressions) but that such a critical feature would be quite insignificant in an overall comparison between a specimen *Q* and a template. It is evident, then, that figures are not always categorized on the basis of their overall, global properties (*op. cit.*, p. 64).

This observation leads directly into consideration of the rival theoretical approach to pattern recognition, 'feature-analysis', where such details as the downward oblique line segment in a '*Q*' *are* critical for recognition. In a computer program like that suggested by Selfridge ([1959]), it is supposed *any* pattern can in principle be recognised, simply by including enough feature analysers in various weighted combinations. Moreover, the machine can improve its performance according to recognition successes or failures, by increasing or decreasing weights assigned to the different features it analyses; and since there is no limited set of patterns it is designed to recognise, a machine with this program could conceivably recognise ill-defined patterns (Neisser, *op. cit.* p. 76).

Is this the way we recognise patterns? Psychologists who prefer feature analysers to templates point to the support this approach gets from cell recordings and other studies done in connection with the visual systems of the octopus (Sutherland [1957]), the frog (Lettvin, Maturana, McCulloch and Pitts [1959]), and the cat (Hubert and Weisel [1959], [1962]). But these are not conclusive: the octopus and the frog respond to very limited ranges of visual stimuli, and even the cat comes nowhere near human performance. The presence of parallel analysers in our own visual cortex, for detecting specifically oriented line segments, would not mean we recognise things the same way cats do, but at most that we detect line orientation the way they do.[1] The dispute over whether feature analysers like 'horizontality' and 'verticality' are innate to the organism (Sutherland [1959]) or developed by experience (Uhr [1963]) is irrelevant to the problem of simulating human pattern recognition if, in fact, such analysers contribute little to our recognition processes. So too would be the point emphasised by Gyr, Brown, Willey, and Zivian ([1966]) that computer simulations involve passive perception of form, while among organisms this is characteristically an active process.

Consider the following rather typical example of reasoning about this problem. Hand-printed letters drawn from the same pool of characters used by Neisser and Weene ([1960]) to test human pattern recognition were presented to a general-purpose computer with a system of about thirty complex feature analysers. After several hundred exposures in which the program could 'learn' to improve performance, unfamiliar letters were presented and identified with close to 90 per cent accuracy (Doyle [1960]). Now that the computer program successfully identified 90 per cent of the trial letters exposed to it, and that it did this by feature analysis of these hand-printed, often sloppy, characters is beyond question. It is also unquestioned that human subjects identify the same letters successfully, usually with 100 per cent accuracy. Does it follow that such computer programs are nearing human capability in this kind of pattern recognition? The whole logic of computer *simulation* of human learning processes strains towards that conclusion.

But in fact it does not follow at all. For all we know, human brains and feature-analysing computer programs may achieve the same, or nearly the same, end in such a simple pattern recognition task by entirely different means. In that case it would be wrong to say the computer is nearing *human* capability in pattern recognition.[2] Rather it is achieving *a* capability for the same task which has almost the same result, but tells us nothing about how humans achieve theirs. And if the latter is really very different, then it might be a costly mistake to suppose we could improve upon the former, until it does—more efficiently and tirelessly—what we do when we sort letters, translate from one language to another, or read a newspaper: not to mention separating sheep from goats and hailing a friend in a crowd.

---

[1] In fact cats appear to depend much more than we do on feature analysis for recognition (see section 3 *below*).

[2] Dogs swim, we swim, fishes swim. Nothing seems gained by saying dogs have 20 per cent of human swimming capability, or men 20 per cent of piscine swimming capability. It is too obvious these animals do the same thing in different ways. We do say, 'He swims like a fish', but all we mean by this is that he swims with great ease for a human.

2 MODES OF PERCEPTION IN LEFT AND RIGHT CEREBRAL HEMISPHERES

It has been known for a long time that damage to the left cerebral hemisphere in right-handed adults often causes loss of speech functions (Broca [1865]). More recently it has been established that lesions of the right hemisphere give rise to disturbances of spatial perception and construction (McFie, Piercy and Zangwill [1950], Hécaen, Ajuriaguerra and Massonet [1951]). Extensive studies of brain wounds incurred in combat confirm this asymmetry: men with right hemisphere damage suffer perceptual impairment without significant intellectual handicap, while those with left-sided cerebral damage show a decline in verbal intelligence, but little or no loss of perceptual judgment in tasks like tactile size discriminations (Weinstein [1962]).

Dependence on the mute hemisphere for form recognition is strikingly illustrated in the case of lesions there causing prosopagnosia, or facial agnosia. In this condition the patient perceives faces but simply fails to recognise them. Here are typical statements:

> If somebody speaks to me, I recognise him from his voice, but his face says nothing to me. If I do not remember the dress she [his wife] is wearing, she could pass right by me and I would not recognise her.
> This man [his brother in a picture] has the same glasses as my brother. . . . No, I don't think [it is him].
> I think it [a picture of her husband] is my husband; looking at his tie, I should say it is my husband, and the face is also like that of my husband.
> [Looking at his wedding picture]: Two people . . . one of them could be my wife, because of the silhouette . . . if it is my wife, the other could be me.

Hécaen and Angelergues ([1962]) provide still other examples: patients who could not distinguish a dog from a cat, or a human being from a long-haired dog, and were able to identify a rabbit only by noting it had long ears.

> There are patients who cannot identify themselves in a mirror. Walking in a shop, Charcot's patient saw a man approaching him and stepped aside to let him pass; the 'man' proved to be his own mirrored image. Bodamer reports concerning his first patient that when asked to look into the mirror, he first

interpreted the image as a picture. Finally he stated that he saw a face which he inferred was his own, but that he really did not recognise it (p. 93). A variety of cues—size, gait, clothes, voice, surroundings—are utilised to achieve identification. Thus, a patient may no longer recognise his physician if the latter takes off his glasses or changes his coat (p. 94).

But in perceptual deficits arising from brain damage there is always the suspicion that the other hemisphere is involved as well. Studies of patients who have had the cerebral commissures surgically disconnected, for control of intractable epilepsy, provide better clues to recognitive abilities. With these patients it is possible to assess the mental capacities of each hemisphere independently for the same perceptual task, by restricting sensory input to one visual field or one side of the body: in the absence of commissural connections what is perceived in one hemisphere is not relayed to the other.

Quite early in the present series of tests it was found that for right handers there is strong lateralisation of speech and calculation in the disconnected left hemisphere, and a specialisation for visuospatial construction in the right, speechless hemisphere (Sperry, Gazzaniga and Bogen [1969]). But I shall concentrate on a very recent experiment involving projection of chimeric figures to the separated hemispheres (Levy, Trevarthen and Sperry [1972]).

In this experiment, different images were simultaneously projected very briefly (150 msec.) into the disconnected hemispheres of four subjects. Each was then asked to identify what he or she had seen by pointing to a corresponding figure among an array of several with one or the other hand, or, subsequently, to state what was seen verbally. Since speech is confined to the left hemisphere in these right-handed patients, the test provided a convenient way of determining that hemisphere's perceptual abilities where a purely verbal response was called for. In the case of a manual response, a significantly higher score for correct identification of what appeared in the right, mute hemisphere rather than what appeared in the left would indicate that the former is better able to make the visual match.

The results confirmed this expectation. In the test for facial recognition, where the correct face had to be pointed to out of three faces, the right hemisphere face was selected 89 per cent of the time when using the right hand. (Since each hand is normally controlled by the contralateral hemisphere, this indicates also that the right hemisphere is able to dominate voluntary motor response for a recognition task.) When it came to a verbal response, what had been seen in the speechless hemisphere was of course named only at chance level. More interestingly, the left hemisphere identification was correct only 49 per cent of the time: indicating a decided inferiority to the right hemisphere.

In a second test requiring the apprehension of complex unfamiliar forms, rather like antler patterns, very similar results were obtained: 79 per cent right hemisphere preference with the left hand, 70 per cent with the right hand, only 51 per cent correct identification of the patterns by vocal naming (hence of what was seen in the left hemisphere). A much higher error rate occurred, presumably because the patterns were harder both to recognise and to name.

Quite familiar objects—drawings of a rose, an eye, and a bee—were projected to the hemispheres in a third test. This time the right hemisphere image dominated 94 per cent of the time with either hand pointing, and the left hemisphere image was named correctly 63 per cent of the time a verbal response was called for.

Finally the hemispheres were tested for recognising chain patterns consisting of various arrangements of crosses and squares in vertical columns of three. Here the right hemisphere preference for left and right hand pointing dropped, indicating that these simple patterns were just as difficult for the mute hemisphere to recognise as the 'antlers' had been. By contrast, the left hemisphere found it three times as easy to identify the chain patterns than the antler-like ones. What accounts for this dramatic reversal? The authors find an explanation in the fact that chain patterns lend themselves to piecemeal, sequential analysis ('one cross, a square and a cross'; 'two squares over a cross'; 'a cross followed by two squares', *etc.*), which is just the sort of thing the left hemisphere does well.

Conversely, the right hemisphere showed a clear superiority for recognition of any visual stimulus which is resistant to analytical verbal description. Going back to the facial recognition test, each of the three faces was given a name. The left hemisphere then learned—somewhat painfully—to associate a name with each face by assigning some verbal label to a distinctive feature of it: 'Dick has glasses', 'Paul has a moustache', 'Bob has nothing' (meaning neither). I quote:

> It was evident in the hesitancy and incidental comments of the subjects as well, that the left hemisphere found this task extremely difficult and was inclined to describe the distinctive features of the right field face [the one it could see] instead of naming it as a unit (p. 66). With the major [left] hemisphere in command the recognition process tended to be more analytical. The major hemisphere kept searching for distinctive subordinate features that could be handled verbally. It seemed unable, for example, to recognise or remember a face as a whole, needing instead to pick out some analytical detail to which a verbal label could be attached. All four of the present tests confirmed in different ways the presence of a fundamental difference in the way the right and left hemispheres

perceive things. . . . It appears from the results that the left hemisphere became dominant when some kind of verbal and/or conceptual, symbolic transformation was required. When immediate recognition and memory of visual shape was the only requirement, even though a stimulus possessed a well-known name, the right hemisphere dominated (p. 74). Our results conform with the assumption that visual retention for direct matching encoding is done by the right hemisphere, and verbal encoding by the left and indicate further that in the absence of biasing influences, the primary tendency is for all visual stimuli to be perceived in the minor [right] hemisphere (p. 75).

3 IDENTIFICATION AND RECOGNITION

I think it is clear from the above data that what the human left cerebral hemisphere (in right-handers) typically does when processing a visual form stimulus is not to recognise it, but to *identify* it. By this I mean it infers from some features or properties of the stimulus *what kind of a thing* it has seen. 'That must be a rose—it has no eyelashes', 'Those antlers look too far apart to be the moose's', 'That couldn't have been Dick because he wasn't wearing glasses, nor Paul because I didn't see a moustache, so . . .': these would appear to the strategies our speaking hemisphere employs. Now identification, while it is a way of coming to know what a visually-presented object is, is not the same as recognition. Suppose I am taking an eye examination and the letter at the top of the chart is an *E*. If I see it easily I say without hesitation and without any introspectively noticeable reasoning process, '*E*'. But if it is blurry to me and I'm not sure what the letter is I might succeed in another way. I might note first that it has a vertical bar in it, and what look like three horizontal bars adjoining the vertical bar to the right; then I might rapidly run through mental pictures of capital letters in the alphabet and stop at *E*. But to be sure it is an *E*, not an *F*, I might have to go back and count the horizontal bars: one, two, three, 'Yes, it is an *E*'. What I have done is to *establish that* I saw an *E*, not to recognise what I saw for an *E*.[1]

Now apply this to the problem of pattern recognition in computers. The most successful programs developed so far embody feature analysing, and this is precisely the strategem—along with inference from what is analysed—employed by the perceptually inferior left half brain. What such programs amount to, if my analysis is correct, is not recognition but identification of the visual stimulus. However this does not matter. Optimally one can hope to develop machine programs that will identify forms as precisely and rapidly as humans recognise them, in which case no one will care what we call what they are doing. But the trouble goes deeper than that. If one takes the brain studies seriously, there is just no chance that identification can successfully substitute for recognition. Gibson, Filbey and Gazzaniga [1970] have shown that even in the undamaged, commissurally-intact human brain, patterns presented to the left cerebral hemisphere are sent over to the right half brain for decoding and relaying back before a verbal name is applied to it. It is extremely unlikely that, if identification were a serious rival to recognition as a means of processing visual information, we should have evolved such interhemispheric dependency. Indeed the evidence from perceptual deficits in otherwise normal left-handed humans, presumed to have some bilateralisation of speech, indicates that the asymmetry of our cerebral functions has evolutionary advantage (Levy [1969]). Apparently when our hominid ancestors developed verbal-analytic-symbolic capacities these were confined to the left hemisphere because their presence in the same hemisphere with Gestalt-synthetic-perceptual abilities is actually antagonistic to the latter (Levy, Trevarthen and Sperry [1972]). But a computer, if it is a computer, cannot operate on other than analytic principles. So it appears the basic *modus operandi* of computers, and hence of all future computers as well, is unsuited to recognition.[1]

---

[1] Sayre ([1965], chapter six) distinguishes recognition from identification in several ways. He does not appear to have considered that feature analysing involves the latter rather than the former.

The point can be seen another way. Suppose you want to meet a friend coming into the airport but are detained, so you ask your wife to meet him. As it happens, your wife does not know him at all. So you describe him to her. When you do this you are thinking of his face as it is visually encoded in the right hemisphere, but all you can *say* about it is that it has certain features ('Roundish face, smooth complexion, pencil-line moustache—if he hasn't shaved it—balding but not grey hair', *etc.*), hoping these will be distinctive enough for her to pick him out among the other arriving male passengers about his age. You are not really preparing her to recognise him (as you might if you were a police artist and could sketch the face), but to recognise *that* it is your friend when she sees him: in other words, to identify him.[2] And this is all you can do verbally, by analysing with your left hemisphere the right hemisphere visual stimulus relayed to it. But computers are in your wife's situation. They can only work with what we give them as visual decoding procedures, and we can only give them information-processing strategies modelled on the verbal hemisphere's quite inadequate ones.

---

[1] Nothing said here rules out designing a machine modelled on the neurological structure of the mute hemisphere, so it does the same thing (whatever that is) our right half brain does when we recognise forms. But it would not be a computer-based machine.

[2] We do sometimes say, 'I recognised him from your description'. If I am right this should be understood to mean, 'I identified him from the features you described'. Thereafter, of course, the speaker is in a position to recognise this person.

The mathematician Bellman ([1968]) gives an excellent example of this limitation in a practical context.

> What is involved in examination for cancer of the cervix? A tissue smear is examined through the microscope or with the aid of a blown-up photograph. An expert looks at the various cell configurations and makes a judgment. Cancer is either present or not present. The problem, then, is one of pattern recognition. Presumably, we should be able to computerise this task in a straightforward fashion. The sensible approach is to take several hundred thousand tissue smears, some cancerous and some not, and use a group of experts to provide a concensus concerning the features that impel them to a decision one way or the other. Unfortunately . . . the pathologist can recognise the condition when it arises, but he cannot tell in complete detail in advance how he goes about the decision process. Several groups at IBM and other research installations have worked on this problem for years, but so far the results apparently have been nil (pp. 108–9).

Generalising now about the problem, he goes on to make the following observations, which will serve to close this section of my paper:

> The interesting fact that comes through as soon as these processes of pattern recognition are analysed is that we can perform all sorts of intricate operations that we really don't understand. In actuality, we observe the opposite of the standard dictum of the high school teacher, 'If you really understand it, you can explain it in words'. It seems that there are no human actions of any subtlety, elegance, or importance that we totally comprehend. There is no way to tell someone how to develop a good backhand in tennis, how to play chess, how to compose music, how to do research, *etc.*, *etc.*, *etc.* We can demonstrate some of the simpler aspects of the entire process, we can provide good and bad examples, and we can correct some obvious faults; but we cannot dissect the activity in a step-by-step fashion. This is the principal reason why the computer is limited in its abilities at present. Furthermore, it is not clear that the situation will ever significantly improve (*ibid.*).

4 CHESS PLAYING MACHINES

On the question, what are the limits of artificial intelligence research?, Feigenbaum and Feldman ([1963]) have given this ringing answer:

> What is important is that we continue to strike out in the direction of the milestone that represents the capabilities of human intelligence. Is there any reason to suppose that we shall never get there? None whatever. Not a single piece of evidence, no logical argument, no proof or theorem has ever been advanced which demonstrates an insurmountable hurdle along the continuum . . . (p. 8).

Many writers take chess playing to be the paradigmatic task for artificial intelligence. For example, Newell, Shaw and Simon [1963]:

> Chess is the intellectual game *par excellence*. Without a chance device to obscure the contest, it pits two intellects against each other in a situation so complex that

> neither can hope to understand it completely, but sufficiently amenable to analysis that each can hope to outthink his opponent. The game is sufficiently deep and subtle in its implications to have supported the rise of professional players, and to have allowed a deepening analysis through 200 years of intensive study and play without becoming exhausted or barren. Such characteristics mark chess as a natural area for attempts at mechanisation. If one could devise a successful chess machine, one would seem to have penetrated to the core of human intellectual endeavor (p. 39).

I am not interested in making sweeping judgments about the potentialities of machine intelligence. Obviously there is more than one way many problem-solving tasks may be successfully completed. No one would say, for example, that mechanical superiority is demonstrated by a computer program that plays unbeatable tick tack toe. In such a simple game, where all one needs to do to draw is to prevent one's opponent from writing down three noughts or crosses in a row on a grid of nine squares, it would be surprising if the trained program ever lost. And there must be many other, more complex, games in which the superior speed[1] and faultless memory storage of a learning machine can overwhelm the human opponent—what might be called the Fuzzy Wuzzy Effect. However the really interesting 'game environments' are those in which these quantitative advantages for the machine are insufficient to secure wins against skilled human players: not because this demonstrates permanent human superiority over machines (who knows the limitations of machines?), but because that fact might tell us something about differential information-processing strategies in human beings.

Consider for example the strikingly disparate results obtained by pitting learning machine programs against skilled players of draughts ('checkers' in North America) on the one hand and of chess on the other. As long ago as 1962 the Samuel draughts program beat a former Connecticut state champion, R. W. Nealey, who said he had not had such competition in the end-game from a human opponent since the last time he lost a game, eight years before (Samuel [1963], p. 104). By contrast, Newell and Simon's ([1958]) prediction that a digital computer would be chess champion of the world by 1968, thought conservative by some machine enthusiasts at the time (Armer [1963], p. 405), now seems hollow. To date no chess program plays at better than sound novice level: between 1,200 and 1,500 on the International Chess Federation Scale, where a master is rated at 2,200 and the reigning world champion, Fischer, has a rating of 2,785. What accounts for this difference? The short answer, that chess is more complex than draughts because it has many more possible continuations—10^{125} compared to 10^{40}—is misleading. Both numbers are so huge they eclipse the ability of either a man or a machine to examine all possible continuations: which is why, of course, rote playing can be defeated so easily.[1] It is true that a computer program examines many more moves than a human player. Samuel's program, working with 6,800 instructions, examined a 'tree' of moves and possible replies to a depth of 20 levels. Current chess programs of the Shannon 'look-ahead' type routinely consider half a million hypothetical moves, penetrating to a depth of five or more half moves.[2] The sophisticated 'advice-taking' program developed by Zobrist and Carlson [1973] cuts this search to about 10,000 hypothetical moves (15 seconds for the computer), by generating from 1,000 to 3,000 'snapshots' or stored patterns describing chess combinations that occur frequently in human play, and correlating these with the board situation. A human chess master, however, typically considers less than 100 possibilities (de Groot [1966]). Obviously the machine has to search farther because it has not developed all the heuristic devices by which a human player automatically discounts unprofitable or sterile continuations without exploring them to any depth. But this would be true of both games, and still leaves unexplained the relative success of draughts programs compared to chess programs.

---

[1] Counting the time it takes a neuron in the human brain to recover its prestimulated state, so as to be able to act on a second stimulus, computer components have a speed advantage on the order of 10^5, even in the older generation of computers.

Perhaps the correct answer lies not in the greater arithmetical complexity of chess, but in its greater geometric complexity. Draughts has only one kind of piece, which moves in one direction only (until crowned, when it can move backwards as well as forwards), and captures are forced. These features give the game a static quality: it is eminently suited for sequential, step-by-step analysis. Chess has six kinds of pieces, all of which (except pawns) move differently in any direction, and captures are never forced unless it is to stave off checkmate. Such features assure for the game a dynamic character which is, I think, highly resistant to linear reductive analysis, except in straightforward end-game play. Admittedly this is a paradoxical claim in view of the voluminous chess analysis which has been and is being published all the time. But we must not be deceived by this. It may be that good play is not achieved by analytic procedures, though after-the-fact analysis shows it to be good play. Consider the following opening:

---

[1] Samuel [1963] estimated that if a machine program could examine three draughts moves per millimicrosecond it would take 10^{21} centuries to canvass 10^{40} possible moves. Zobrist and Carlson [1973] estimate that, given 10^{125} possible continuations, there are 10^{43} possible board situations in chess, with only 10^{16} microseconds available in a century to look at them. Since men have certainly played less than 10^{15} chess games, no more than an infinitesimal fraction of chess positions have ever been seen.

[2] A 'whole' move counts as one's own move *and* the opponent's probable reply.

	White	*Black*
1.	*P–K4*	*P–K4*
2.	*P–Q4*	*P–Q3*
3.	*N–KB3*	*N–Q2*
4.	*B–QB4*	*N(N1)–B3*
5.	*B(B1)–N5*	*B–K2*
6.	*Q–Q3*	o–o
7.	o–o	*P(KR2)–R3*
8.	*B–R4*	*P–B3*
9.	*N–B3*	*R–K1*
10.	*B(QB4)–N3*	*N–R4*

In an identical board situation, Newell, Shaw and Simon [1963] provides this as a scrap of White's running comment on his own position:

> Are there any other threats? Black also has a threat of Knight to Bishop 5 threatening the Queen, and also putting more pressure on the King's side because his Queen's Bishop can come over after he moves his Knight at Queen 2; however, that is not the immediate threat. Otherwise, his Pawn at King 4 is threatening my Pawn . . . (p. 40).

Now White may very well *say* this, if invited to speak aloud on his estimate of the board situation. But as any player knows, one does not do a verbal analysis of positions when playing, not even silently to oneself. Rather one *sees* the Black Knight bounding over to Bishop 5, secured there by its Pawn, pressing down as it were on the White Queen. One *sees* (and quickly rejects) the White Pawn at Knight 2 advancing to ward this Black Knight off, but thereby weakening the King's castled defence. And one *sees* instead the White Knight moving to King 2, where it can do the same job better (but at a price). . . . It is true as these authors say that, for example, 'proving theorems and playing chess involve the same problem of reasoning with heuristics that select fruitful paths of exploitation in a space of possibilities that grows exponentially' (p. 50)—*for a machine*. But however humans prove theorems, I am suggesting they do not 'reason out' chess problems at all. What they do is see the flow of pieces over the board in a dynamic way, leading to pressures and exchanges, gains and losses of *tempi* or material or security that will ultimately spell victory or defeat. A computer chess program cannot do this: it must substitute for the perceptual process an extensive search procedure which, depending on the sophistication of its heuristic devices, may or may not allow it to achieve a similar game result.

Simon and Barenfeld ([1969]), convinced by studies of eye movements and verbal commentaries of chess subjects that during the first 5 or 10 seconds of studying a board situation the human player is more concerned with extracting information about the position than with exploring sequences of moves, have designed a computer program called PERCEIVER to duplicate this initial information-gathering phase. That PERCEIVER does the same thing as a human player is evidenced, they think, by the program's attending to the same critical pieces on the board (p. 477). But of course it does not follow that if PERCEIVER and the human player 'fixate' on the same pieces in the same sequence they see them the same way. De Groot ([1966]) has shown that within 5 seconds a master player can get more relevant information about the position than a lesser player can accumulate in a quarter of an hour of analysis; and that a 5-second exposure to a complicated middle game position is sufficient for the master to be able to reproduce it with very few errors, if any, while even a player of expert rank cannot do this. Yet chess masters are *not generally* superior to control subjects in visual memory achievement. De Groot concludes:

> By their *specialty*, therefore, masters must have a superior coding system immediately available. What they 'see' must be something quite different from what lesser players see (p. 35).

Yet when the pieces are placed at random on the board, so they no longer form a meaningful *chess pattern*, master players are no better than weak players in reproducing the position after a 5-second exposure. What would appear to be crucial to master-level play is not superior *general* knowledge of chess possibilities and probabilities, but a way of organising the visual information into configurations that lay bare the dynamic interrelationships of the pieces. And this is a Gestalt-synthetic process, not a sequential, analytic one.

Zobrist and Carlson ([1973]) have noted the fact that no amount of studying games by masters and reading their analyses of these is sufficient to convert an average player into a player of the first rank.[1] As they suggest:

> It is possible that a significant portion of human chess technique cannot be expressed in words. For example, a grand master might be quite unable to explain the reasoning behind a particularly brilliant move. It is not just a question of whether one can have thoughts without words but of whether one can have thoughts of much greater content than the words one can find to express them. It is possible that chess thought depends heavily on spatial perception and that the perceptual processes involved are so subtle and rapid that only the final outcome reaches conscious expression. If so, there may be no adequate language for conveying chess knowledge to a computer (p. 101).

---

[1] One might add to this that while professional players devote most of their time to the game, there are enough examples of chess prodigies (Reshevsky, Fischer, the Byrne brothers), lazy masters (Capablanca), and quite isolated geniuses (Morphy) to support the view that one does not *become* a master-level player by working at the game.

I am inclined to believe this is the case, and furthermore that chess perception has a neurophysiological basis in the synthesising information-processing activities of the right cerebral hemisphere. The left half brain, if I am correct, makes a minimal contribution to chess playing. It does the after-the-fact analysing and explaining, it makes the running verbal commentaries on what the right hemisphere is seeing, it writes the chess books and reads them. But all this verbal encoding and analytical dissecting is, if I am right, completely secondary to the visual encoding and Gestalt-processing done by its mute cerebral companion. The relationship can be understood on analogy with a choreographer's skills. He has to write down in a conventional notation system where each dancer on the stage is to step, what movements they are to execute, *etc.*—typically left hemisphere activity. But the visualisation of the dance sequence, the flow of bodily movements, the dancers' gestures, stance, postures, and so on—all of these processes would seem to originate in and be controlled by the nonspeaking hemisphere. Without verbal-graphic-analytic skills there cannot be choreography. But without visual spatial perception there cannot be an art of dance.

Some empirical predictions are entailed in these observations, though it is difficult to see how they would be substantiated conclusively. It should be the case if I am correct that, for example, left-handers with bilateralised speech functions would be weaker players than dextrals with the same chess training. A right-handed master rendered aphasic by gross lesions in the left hemisphere should be unaffected in his chess playing, provided the mute hemisphere is not also involved. Differential brain wave activity as measured by *EEG* recordings of silent chess players would provide partial confirmation, or disconfirmation, as the case may be.[1] If it were medically justifiable, Amytal injections in one or the other carotid artery, temporarily depressing the ipsilateral cerebral hemisphere, should make considerable difference to a master's reproduction of a chess position after brief exposure to it. But I have thought of a more dramatic test, were it technically feasible.

In his story *Schachnovelle*, Zweig ([1945]) portrays a chessmaster going mad playing his greatest game against himself. Experienced players will find this idea quite unrealistic, since in such a game there would be no surprises: White always knows what Black is up to, and vice versa. However a cerebrally commissurotomised subject with pre-operative chess skills could, under certain conditions, play 'against himself'; meaning by this that left and right hemispheres might be induced to play opposite sides of the board without, in the absence of commissural connections, either hemisphere knowing what its opponent is up to. Various considerations weigh against the likelihood of devising an experiment that would allow this to happen. None of the split-brain patients tested so far were players (Sperry [1972]); often they are brain-damaged and of substandard *IQ*. New operative techniques, leaving the posterior one-third of the corpus callosum intact, have alleviated epileptic seizures in two patients without preventing interhemispheric integration (Gordon, Bogen and Sperry [1971]); if this becomes common practice, we may never have a split-brain chess subject. Finally, it is extremely difficult to lateralise information exclusively to one or the other hemisphere (Gazzaniga [1972]), and there is a danger of causing perplexity and confusion to the patient if testing leads to simultaneous double responses (Levy, Trevarthen and Sperry [1972]). However, supposing a chess playing patient with complete sectioning of the neocommissures is available some day, here is how one might attempt to determine differential contributions of the cerebral hemispheres to chess without provoking unnecessary traumas.

---

[1] A higher right/left ratio of power occurs when a subject performs verbal tasks like mentally composing a letter than when performing a spatial task like choosing mentally which of several assembled figures could be constructed from a cut-up paper. Greater output, mostly in the alpha band, is interpreted as indicating *less* hemispheric engagement in the particular cognitive task (see Galin and Ornstein [1972]).

The subject is seated before a table with the chessboard and pieces in starting position presented laterally to view, say with White pieces to his right and Black to his left. Below this table is a shelf with a divider in the middle, forming two compartments into which he is asked to place his hands. Raised plastic letters and numbers 1⅛″ high are arranged there in vertical columns, as follows:

P	1	*P*	1
R	2	*R*	2
N	3	*N*	3
B	4	*B*	4
Q	5	*Q*	5
K	6	*K*	6
	7		7
	8		8

The subject is told he is going to play chess silently, signalling his move by finding with either hand the letters and number necessary for a standard notation and pressing firmly on the ones he wants, *e.g.* '*P–K*4'. Now the subject is asked to fixate on a point in the middle of a screen above the chess table, where he will be able to read which pieces are his. Having confirmed fixation, the examiner flashes tachistoscopically (1/10 of a second or less) the word WHITE on the right side of the screen (seen only in the left hemisphere) and BLACK on the left side (seen only by the right hemisphere). The examiner moves all the pieces himself.

I believe that in these circumstances the subject will play both Black and White pieces successively, *i.e.* that the hemispheres, each thinking it is playing against the examiner, will play against each other. I do not think, given the normally weak ipsilateral projection of the tactile system (Levy, Nebes and Sperry [1971]), either hemisphere would feel the same-sided fingers selecting and pressing letters or numbers when this particular hemisphere is waiting for 'the examiner's' next move. If the mute hemisphere preferred the dominant hand to signal its move, the speech hemisphere might indeed become suspicious, since it could feel those contralateral fingers selecting keys. But if the subject is an avid enough player he may not notice this, any more than most players notice they are drumming their fingers on the table top when engrossed in a game. The really interesting question is whether the speech hemisphere, left to its own resources, has the means of visually processing significant chess patterns, and thus putting up a good fight. If long-term memory is unaffected by the surgery, as appears is the case (Sperry [1968]), neither hemisphere will have had a greater experience of openings, tactical combinations or strategic lines of play than the other, so it should be an eminently fair contest. After all, they played all the same games together before.

ROLAND PUCCETTI
Dalhousie University

REFERENCES

ARMER, P. [1963]: 'Attitudes towards Intelligent Machines', in E. A. Feigenbaum and J. Feldman (*eds.*): *Computers and Thought*, pp. 389–405.
BELLMAN, R. [1968]: *Some Vistas of Modern Mathematics.*
BROCA, P. [1865]: 'Sur la Faculté du Langage Articulé', *Bulletin Soc. d'Anthropologie*, **6.**
DE GROOT, A. D. [1966]: 'Perception and Memory Versus Thought: Some Old Ideas and Recent Findings', in B. Kleinmuntz (*ed.*): *Problem Solving*, pp. 19–50.
DOYLE, W. [1960]: 'Recognition of Sloppy, Hand-Printed Characters', *Proceedings of the Western Joint Computer Conference* (reported in Neisser [1967], p. 76).
FEIGENBAUM, E. A. and FELDMAN, J. [1963]: *Computers and Thought.*
GALIN, D. and ORNSTEIN, R. [1972]: 'Lateral Specialization of Cognitive Mode: an EEG Study', *Psychophysiology*, **9**, pp. 412–18.
GAZZANIGA, M. S. [1972]: Personal Interview with the author.
GIBSON, A. R., FILBEY, R. A. and GAZZANIGA, M. S. [1970]: 'Hemispheric Differences as Reflected by Reaction Time', *Federation Proceedings*, **29**, p. 658.
GORDON, H. W., BOGEN, J. E. and SPERRY, R. W. [1971]: 'Absence of Deconnexion Syndrome in Two Patients with Partial Section of the Neocommissures', *Brain*, **94**, pp. 327–36.
GUNDERSON, K. [1971]: *Mentality and Machines.*
GYR, J. W., BROWN, J. S., WILLEY, R. and ZIVIAN, A. [1966]: 'Computer Simulation and Psychological Theories of Perception', *Psychological Bulletin*, **65**, pp. 174–92.
HEBB, D. O. [1949]: *The Organization of Behavior.*

HÉCAEN, H., AJURIAGUERRA, J. and MASSONNET, J. [1951]: 'Les Troubles Visio-constructifs par Lésion Parieto-occipitale Droite', *Encéphale*, **1**, pp. 122–79.

HÉCAEN, H. and ANGELERGUES, R. [1962]: 'Agnosia for Faces (Prosopagnosia)', *Archives of Neurology*, **7**, pp. 92–100.

HÖFFDING, H. [1891]: *Outlines of Psychology.*

HUBEL, D. H. and WIESEL, T. N. [1959]: 'Receptive Fields of Single Neurones in the Cat's Striate Cortex', *Journal of Physiology*, **148**, pp. 574–91.

HUBEL, D. H. and WIESEL, T. N. [1962]: 'Receptive Fields, Binocular Interaction, and Functional Architecture in the Cat's Visual Cortex,' *Journal of Physiology*, **160**, pp. 106–54.

LETTVIN, J. Y., MATURANA, H. R., MCCULLOCH, W. S. and PITTS, W. H. [1959]: 'What the Frog's Eye tells the Frog's Brain', *Proceedings of the Institute of Radio Engineers*, **47**, pp. 1940–51.

LEVY, J. [1969]: 'Possible Basis for the Evolution of Lateral Specialization of the Human Brain', *Nature*, **224**, pp. 614–15.

LEVY, J., NEBES, R. D. and SPERRY, R. W. [1971]: 'Expressive Language in the Surgically Separated Minor Hemisphere', *Cortex*, **7**, pp. 49–58.

LEVY, J., TREVARTHEN, C. and SPERRY, R. W. [1972]: 'Perception of Bilateral Chimeric Figures Following Hemispheric Deconnexion', *Brain*, **95**, pp. 61–78.

MCFIE, J., PIERCY, M. F. and ZANGWILL, O. L. [1950]: 'Visual Spatial Agnosia Associated with Lesions of the Right Cerebral Hemisphere', *Brain*, **73**, pp. 167–90.

NEISSER, U. [1967]: *Cognitive Psychology.*

NEISSER, U. and WEENE, P. [1960]: 'A Note on Human Recognition of Hand-Printed Characters', *Information and Control*, **3**, pp. 191–6.

NEWELL, A., SHAW, J. C. and SIMON, H. A. [1963]: 'Chess-playing Programs and the Problem of Complexity', reprinted in E. A. Feigenbaum and J. Feldman (*eds.*): *Computers and Thought*, pp. 39–70.

NEWELL, A. and SIMON, H. A. [1958]: 'Heuristic Problem-solving: the Next Advance in Operations Research', *Operations Research*, **6**, I.

SAMUEL, A. L. [1963]: 'Some Studies in Machine Learning Using the Game of Checkers', reprinted in E. A. Feigenbaum and J. Feldman (*eds.*): *Computers and Thought*, pp. 71–105.

SAYRE, K. M. [1965]: *Recognition: A Study in the Philosophy of Artificial Intelligence.*

SELFRIDGE, O. G. [1959]: 'Pandemonium: A Paradigm for Learning', in *The Mechanization of Thought Processes.*

SIMON, H. A. and BARENFELD, M. [1969]: 'Information-Processing Analysis of Perceptual Processes in Problem Solving', *Psychological Review*, **76**, pp. 473–83.

SUTHERLAND, N. S. [1957]: 'Visual Discrimination of Shape by Octopus', *British Journal of Psychology*, **48**, pp. 55–70.

SUTHERLAND, N. S. [1959]: 'Stimulus Analyzing Mechanisms', in *The Mechanisation of Thought Processes.*

SPERRY, R. W. [1968]: 'Hemisphere Deconnection and Unity of Conscious Awareness', *American Psychologist*, **23**, pp. 723–33.

SPERRY, R. W. [1972]: Personal Interview with the author.

SPERRY, R. W., GAZZANIGA, M. S. and BOGEN, J. E. [1969]: 'Interhemispheric Relationships: the Neocortical Commissures; Syndromes of Hemisphere Disconnection', in P. J. Vinken and G. W. Bruyn (*eds.*): *Handbook of Clinical Neurology*, **4**, pp. 273–90.

UHR, L. [1963]: '"Pattern Recognition" Computers as Models for Form Perception', *Psychological Bulletin*, **60**, pp. 40–73.

WEINSTEIN, S. [1962]: 'Differences in Effects of Brain Wounds Implicating Right or Left Hemispheres', in V. B. Mountcastle (*ed.*): *Interhemispheric Relations and Cerebral Dominance*, pp. 159–76.

ZOBRIST, A. L. and CARLSON, F. R. Jr. [1973]: 'An Advice-Taking Chess Computer', *Scientific American*, **228**, pp. 92–105.

ZWEIG, S. [1945]: *Schachnovelle.*

Brit. J. Phil. Sci. **25** (1974), 155–188 *Printed in Great Britain*

Discussions

ON POPPER'S DEFINITIONS OF VERISIMILITUDE[1]

Introduction.
1 *Preliminaries.*
2 *Popper's Logical Definition of Verisimilitude.*
3 *Popper's Probabilistic Definition of Verisimilitude.*
4 *Conclusion.*

Introduction.

Sir Karl Popper's epistemological position is best characterised as an optimistic scepticism. It is a scepticism since it affirms that no non-trivial theory can be justified and that more likely than not all the theories we entertain and use are false. The position is optimistic in contending that in science we nevertheless make progress: that we have a way of improving on our false theories. Progress, however, hardly ever consists in supplanting a false theory by a true one. As a rule, the new theory is also false but somehow less so than its antecedent. Popper's epistemology thus calls for a discriminating approach to false theories: he has to assume that of two false theories, one can be preferable to the other in being 'closer to the truth' or 'more like the truth'.

In an attempt to legitimise this sort of talk Popper has proposed two rigorous definitions of verisimilitude, I shall call them logical and probabilistic. The aim of this note is to show that for simple logical reasons, both are totally inadequate.

In Section 1 are given Popper's definitions of several auxiliary notions. The logical definition of verisimilitude is considered in Section 2. It is demonstrated that on this definition a false theory can never enjoy more verisimilitude than another false theory. The probabilistic definition is dealt with in Section 3. An example of two theories A and B is given such that A is patently closer to the truth than B, yet on Popper's definition A has strictly less verisimilitude than B.

1 *Preliminaries.*

Consider a language having (as is usual) a finite number of primitive descriptive constants. Any finite set of (closed) sentences of the language will be called a *theory*. In what follows, A, B, C, . . . are understood to be arbitrary theories. $Cn(A)$ is the set of theorems of A, *i.e.*, the set of logical consequences of A. Furthermore, let T and F be the set of true and false sentences of the language respectively. Popper has proposed the following definitions.[2]

---

[1] An earlier version of this paper was presented to the Philosophy Seminar of the University of Otago in March 1973. The author benefited from conversations with Sir Karl Popper, Alan Musgrave, and John Harris, and adopted a terminological suggestion made by David Miller.

[2] The latest formulations of these definitions can be found in Professor Popper's [1972]. In what follows, all page references are to this book.

Definition 1.1.[1] The *truth content* A_T *of* A is $Cn(A) \cap T$.

Definition 1.2.[2] The *relative content* A, B *of* A *given* B is $Cn(A \cup B) - Cn(B)$.

Definition 1.3.[3] The *falsity content* A_F *of* A is the relative content of A given A_T, *i.e.*, A, A_T.

Definitions *1.1*, *1.2*, and *1.3* yield

Proposition 1.4. $A_F = Cn(A) \cap F$.[4]

Proof.

$$\begin{aligned} A_F = A, A_T &= Cn(A \cup A_T) - Cn(A_T) && \text{(by } 1.3 \text{ and } 1.2\text{)} \\ &= Cn(A) - A_T && \text{(since } A \cup A_T = A \text{ and } Cn(A_T) = A_T\text{)} \\ &= Cn(A) - (Cn(A) \cap T) && \text{(by } 1.1\text{)} \\ &= Cn(A) \cap F && \text{(since } \overline{T} = F\text{).} \end{aligned}$$

2 *Popper's Logical Definition of Verisimilitude.*

Popper never explicitly states but obviously presupposes:

Definition 2.1. A_T and B_T (or A_F and B_F) are *comparable* just in case one of them is a (proper or improper) subclass of the other.

Now we can state Popper's logical definition of verisimilitude:

Definition 2.2.[5] *A has less verisimilitude than B* just in case (*a*) A_T and A_F are respectively comparable with B_T and B_F, and (*b*) either $A_T \subset B_T$ and $A_F \not\subset B_F$ or $B_T \not\subset A_T$ and $B_F \subset A_F$.

Definitions *2.1* and *2.2* yield immediately

Proposition 2.3. A has less verisimilitude than B just in case either $A_T \subset B_T$ and $B_F \subseteq A_F$ or $A_T \subseteq B_T$ and $B_F \subset A_F$.

Definition *2.2* is inadequate as explication of verisimilitude in view of

Proposition 2.4. If B is false then A does not have less verisimilitude than B.

Proof. Since B is false, there is a false sentence, say f, in $Cn(B)$. First assume

---

[1] This is how the concept of truth content is defined on p. 330. On p. 48 we are given a slightly different definition, whereby the truth content of A is rather $(C(A) \cap T) - L$, where L is the set of tautologies or logically valid sentences. But I take this to be a mere slip, since some statements on the same page are in conflict with this definition. At all events, the difference is marginal and does not affect our ensuing considerations.

[2] This is how the concept of relative content is defined on p. 332. On p. 49 the relative content of A given B is characterised as the class of all sentences deducible from A *with the help of* B. This might be construed as suggesting that the relative content of A given B is simply $Cn(A \cup B)$. However, from several subsequent remarks it transpires that this is not what is intended.

[3] See pp. 49, 51 and 332.

[4] The proposition shows that the definition of the falsity content of A (as the class of false consequences of A), which is considered and rejected on p. 48 is in effect logically equivalent to the definition actually proposed at the bottom of p. 49.

[5] See p. 52. The signs $\subseteq$ and $\subset$ stand for set inclusion and proper set inclusion respectively.

$A_T \subset B_T$. Then there is a sentence, say b, in $B_T - A_T$. But then $(f\,.\,b) \in B_F$. On the other hand, $(f\,.\,b) \notin A_F$, since otherwise, by *1.4* and *1.1*, $b \in A_T$, in contradiction to the choice of b. Thus $B_F \not\subseteq A_F$. Now assume $B_F \subset A_F$. Then there is a sentence, say a, in $A_F - B_F$. But then $(f \supset a) \in A_T$. On the other hand, $(f \supset a) \notin A_F$, since otherwise, by *1.1* and *1.4*, $a \in A_T$, in contradiction to the choice of a. Thus $A_T \not\subseteq B_T$. The Proposition now follows by *2.3*.

To illustrate Proposition *2.4*, let A consist of the sole sentence 'It is now between 9.40 and 9.48' and let B consist of the sole sentence 'It is now between 9.45 and 9.48', where 'between' is understood to exclude the two bounds. Suppose that the actual time is 9.48.[1] Then B is false. Moreover, $A_T \subset B_T$. Yet A does not have less verisimilitude than B on Definition *2.2*. For clearly the (only) member of B is in B_F but not in A_F, thus $B_F \not\subseteq A_F$.[2]

3. *Popper's Probabilistic Definition of Verisimilitude.*

Where A and B are theories, let $p(A)$ be the logical probability of A and $p(A, B)$ the relative logical probability of A given B. Popper has proposed the following definitions.

Definition 3.1.[3] The ***measure*** $ct_T(A)$ ***of the truth content of*** A is $1 - p(A_T)$.

Definition 3.2.[3] The ***measure*** $ct_F(A)$ ***of the falsity content of*** A is $1 - p(A, A_T)$.

Popper's probabilistic explication of truthlikeness is then in terms of ct_T and ct_F. Popper offers, in fact, two alternative explications. They will be spoken of as verisimilitude$_1$ and verisimilitude$_2$. The definitions are as follows.

Definition 3.3.[4] The ***verisimilitude***$_1$ $vs_1(A)$ ***of*** A is $ct_T(A) - ct_F(A)$.

Definition 3.4.[4] The ***verisimilitude***$_2$ $vs_2(A)$ ***of*** A is

$$(ct_T(A) - ct_F(A))/(2 - ct_T(A) - ct_F(A)).$$

Both concepts are drastically at variance with the intuitive notion of closeness to the truth. Preparatory to a justification of this claim I shall introduce several notational conventions and prove an auxiliary proposition.

Let $a, b, \ldots, t, \ldots$ be arbitrary sentences of the language in question. In what follows, a symbol standing for a sentence will also be used to denote the set whose only element is that sentence.

Proposition 3.5. If $T = Cn(t)$ then $a_T = Cn(a \vee t)$.

Proof. Assume $T = Cn(t)$ and consider an arbitrary sentence b. By *1.1*,

---

[1] See p. 56.

[2] In private conversation Professor Popper suggested to the author that things might be remedied if we forgot all about falsity contents and simplified Definition *2.2* to the following: A has less verisimilitude than B just in case $A_T \subset B_T$. It is easy to show, however, that on this definition a false theory A has less verisimilitude than another false theory B just in case $Cn(A) \subset Cn(B)$, *i.e.*, just in case A is a logical consequence of B. If this definition was adequate it would be child's play to increase the verisimilitude of any false theory A: it would suffice to add to A an arbitrary sentence which does not follow from it.

In a personal letter David Miller has informed the author that he independently obtained the results of Section 2.

[3] See pp. 51, 337.

[4] See p. 334.

$b \in a_T$ iff $b \in Cn(a) \cap Cn(t)$. But by propositional logic, $b \in Cn(a) \cap Cn(t)$ iff $b \in Cn(a \vee t)$.

The inadequacy of *3.3* and *3.4* will now be demonstrated on a simple example.

Consider a rudimentary weather-language L containing no predicates and only three primitive sentences, 'it is raining', 'it is windy' and 'it is warm'. Let us abbreviate them respectively as 'p', 'q', and 'r'. Moreover, assume all the three sentences are, as a matter of fact, true. Then, writing t for $p \,.\, q \,.\, r$, we have $T = Cn(t)$.

The eight sentences $p \,.\, q \,.\, r$, $p \,.\, q \,.\, {\sim} r, \ldots, {\sim} p \,.\, {\sim} \,.\, q \,.\, {\sim} r$ will be spoken of as *constituents*. The constituents are mutually incompatible, jointly exhaustive and of equal logical strength. Hence the logical probability of each is 1/8. As well known, every consistent sentence a of L is logically equivalent to a disjunction of constituents the (disjunctive) *normal form* of a. The following four propositions clearly hold of any sentences a and b of L:

3.6. a is compatible with b just in case the normal forms of a and b have a constituent in common.

3.7. a is true just in case a is compatible with t.

3.8. If a is incompatible with b then $p(a \vee b) = p(a) + p(b)$.

3.9. The relative probability $p(a, b)$ of a given b is $p(a \,.\, b)/p(b)$.

Now it is easy to prove

Proposition 3.10. If a is false then $ct_T(a) = (7/8) - p(a)$ and

$$ct_F(a) = 1 - [p(a)/(p(a) + 1/8)].$$

Proof. Let a be false. We have:

$$\begin{aligned} ct_T(a) &= 1 - p(a_T) && \text{(by } \textit{3.1}\text{)} \\ &= 1 - p(a \vee t) && \text{(by } \textit{3.5}\text{)} \\ &= 1 - [p(a) + p(t)] && \text{(by } \textit{3.7} \text{ and } \textit{3.8}\text{),} \end{aligned}$$

$$\begin{aligned} ct_F(a) &= 1 - p(a, a_T) && \text{(by } \textit{3.2}\text{)} \\ &= 1 - p(a, a \vee t) && \text{(by } \textit{3.5}\text{)} \\ &= 1 - [p(a \,.\, (a \vee t))/p(a \vee t)] && \text{(by } \textit{3.9}\text{)} \\ &= 1 - [p(a)/(p(a) + p(t))] && \text{(by } \textit{3.7}, \textit{3.8}, \text{ and propositional logic).} \end{aligned}$$

But $p(t) = 1/8$. Which completes the proof.

From *3.10*, *3.1*, and *3.2* it immediately follows that the values of vs_1 and vs_2 at false sentences of L depend solely on the logical probabilities of the sentences. A little reflection reveals that this fact alone makes vs_1 and vs_2 unfit to explicate the intuitive notion of proximity to the truth. Since surely we want it to be possible for one false theory to be closer to the truth than another false theory despite the two theories having the same logical probability. If Popper's proposals were right then in order to decide which one of two false theories is closer to the truth, no factual knowledge would be required over and above the knowledge that the two theories are indeed false. Which is clearly absurd.

To illustrate this point, let us consider a couple of examples.

The following table gives the values of ct_T, ct_F, vs_1, and vs_2 at some false sentences of L:

	$\sim p \,.\, \sim q$	$p \,.\, q \,.\, \sim r$	$\sim p \,.\, \sim q \,.\, \sim r$
ct_T	5/8	6/8	6/8
ct_F	1/3	1/2	1/2
vs_1	7/8	2/8	2/8
vs_2	21/25	1/3	1/3

Now imagine that Jones and Smith, two prisoners sharing a windowless and air-conditioned cell, are using L to discuss the weather. Jones takes the view that it is a dry, still day, with a low temperature. In other words, Jones's conjecture is $\sim p \,.\, \sim q \,.\, \sim r$. Smith disagrees. Although he also thinks that the temperature is low, he (rightly) insists that it is raining and windy. In other words, Smith's theory is $p \,.\, q \,.\, \sim r$.

It seems hardly deniable that Smith is by far nearer to the truth than Jones. He is admittedly wrong on temperature, but he is dead right as far as rain and wind are concerned. Jones, on the other hand, is wrong on three counts. He could not, in fact, be farther from the truth than he is (without contradicting himself).

Thus one would expect Smith's theory to exceed Jones's in measure of truth content and in verisimilitude. One would also expect Jones's theory to exceed Smith's in measure of falsity content. Yet, as seen in the above table, each of the functions ct_T, ct_F, vs_1, and vs_2 takes the same value at Jones's theory as it does at Smith's.

But Popper's functions vs_1 and vs_2 not only fail to discriminate between theories which, like the two above, are vastly unlike in proximity to the truth. In many cases the functions accord strictly greater verisimilitude to a theory which is patently farther from the truth than another theory.

Let us alter slightly the above example. Imagine that while Smith sticks to his theory $p \,.\, q \,.\, \sim r$, Jones has weakened his claim to $\sim p \,.\, \sim q$. Jones's theory is now marginally better than before: while previously Jones was positively wrong on temperature, this time he withholds judgement. But Jones's new theory is surely not better enough to match, let alone exceed, Smith's in closeness to the truth. Jones's is still one of the lousiest and Smith's one of the best false theories, as false theories go. Smith is only wrong on one count, whereas Jones is wrong on two. One would certainly expect Jones's theory to exceed Smith's in falsity content. Yet, the ct_F of Jones's theory is strictly less than the ct_F of Smith's. One would also expect Smith's theory to have greater verisimilitude than Jones's. Yet, the vs_1 of Smith's theory is strictly less than the vs_1 of Jones's and similarly for vs_2.

4 *Conclusion.*

To do justice to the intuitive notion of truthlikeness one must clearly make it possible for a false theory to be closer to the truth than another false theory of the same logical probability. For a simple language which, like L, is based on propositional logic only, this is easily done. The 'distance' between two constituents can be naturally defined as the number of primitive sentences negated in one of the constituents but not in the other. The verisimilitude of an arbitrary sentence a can then be defined as the arithmetical mean of the distances between the true constituent t and the constituents appearing in the disjunctive normal form of a. It is easily seen that such a definition meets all intuitive requirements.

Things, of course, get vastly more complicated when we turn to the more typical kind of theories, *i.e.*, theories formulated in a first-order language. But the idea underlying the above definition of 'distance' can be carried over to first-order theories if one employs, in lieu of disjunctive normal forms, Hintikka's distributive normal forms for first-order formulas. This, however, is a topic for a separate article.

PAVEL TICHÝ
University of Otago,
Dunedin, New Zealand

REFERENCE

POPPER, K. R. [1972]: *Objective Knowledge.*

POPPER'S DEFINITIONS OF 'VERISIMILITUDE'[1]

1 One of the major problems Popper has attacked is that of finding and making intelligible a coherent view of critical common-sense realism which agrees with the practices of science. He sees science as progressing and finding ever better theories. And since for him the only significant progress would be that of getting closer to the (absolute) truth, he is obviously led to the minor problem of explaining what it would mean (at least in principle) to say that one theory is closer to the truth than another, especially in the case when both theories are false. His attempts at explicating this particular concept usually appear in his writings under the heading of 'verisimilitude'.

For any two theories A and B let us write '$A <_T B$' if intuitively B is closer to the truth than A. (This is not to imply that this intuitive concept has a unique sense and won't someday be found to be ambiguous. But whether this is the case is part of the problem.) Popper has given essentially two different formal definitions of '$<_T$', which Miller[2] calls the qualitative and the quantitative definitions. I will denote the first by '$<^1_T$'. Miller[3] and Tichy[4] have independently shown that neither of the formal definitions is faithful to Popper's intuitive notion of verisimilitude. In particular they prove that if A and B are false theories, then on Popper's qualitative definition neither can be closer to the truth than the other:

(*1*) $$A \nless^1_T B \text{ and } B \nless^1_T A.$$

The main purpose of this article is (*i*) to consider a mathematically more general problem, that of comparing theories relative to an arbitrary comparison theory, (*ii*) to show the philosophical relevance of this generalisation and (*iii*) to shed light on the Miller-Tichy result (*1*) by obtaining some general results of which (*1*) is a special case.

2 When we say that science finds ever better theories or that one theory is better than another, implicit in such a statement is the assumption that there is some criterion of comparison such as elegance, ease of calculation, degree of falsifiability, agreement with a portion of currently accepted background know-

---

[1] I am deeply indebted to Pavel Tichý for bringing to my attention the problem of verisimilitude and his negative results. I am also indebted to P. Tichý, D. Miller and A. Musgrave for criticism of earlier drafts of this article.

[2] Miller [1974].

[3] *Ibid.*

[4] Tichý [1974].

ledge, or agreement with the truth. In this paper I will focus on the last two types of criteria.

By definition a *theory* (over language L) is any set of sentences of L. In this paper let us restrict our consideration just to consistent theories.

Let us write '$A <_C B$' if intuitively theory B agrees better with the theory C than does theory A, where the sentences in C are thought of as assumed truths. I will call any such set C a *comparison theory*. Scientists seem to have a clear intuitive idea of when $A <_C B$, at least in many particular instances of A, B and C. In such cases the comparison theory C consists of a portion of accepted background knowledge, some assumed or alleged facts. For example, scientists agree that Einstein's general theory of relativity (E) agrees better with experimental data than does Newton's classical theory (N). In comparing theories, it is not the 'truth-set' that scientists use, but some set of statements which are assumed true for the purposes of the comparison.

The problem of finding a good definition of '$A <_T B$' may seem to be a purely logical problem, having nothing to do with how scientists compare theories in actual practice. But what criteria should such a definition satisfy? I claim that one thing gained by generalising the problem of defining '$A <_T B$' to that of defining '$A <_C B$' for an arbitrary comparison theory C are some very specific desiderata for any definition of verisimilitude. In particular I propose that any good definition of '$A <_T B$' when naturally generalised to an arbitrary comparison theory C must give a definition of '$A <_C B$' which agrees with the scientist's intuitive notion of '$A <_C B$'. Turning this around I propose that we should look for a formal definition of the scientist's intuitive notion of '$A <_C B$'; and any such definition should be general in form so that it applies to any comparison theory C. Then if we are working in a language with a standard interpretation and a corresponding truth set T, we will have the desired definition of '$A <_T B$' by letting $C = T$.

Finally, if one interprets '$A <_C B$' as saying that theory B is better corroborated than theory A by the comparison set C, then the problem of defining 'verisimilitude' can be considered just a special case of the problem of defining 'corroboration'.

3 In this section let us review the literature and notation for Popper's qualitative definition of '$A <_T B$'. Hence we will temporarily restrict the comparison theory to T. Assume we are working with theories over a fixed language L which has a standard interpretation so that T (the set of all true sentences) and F (the set of all false sentences) are well-defined.

Central to both his qualitative and his quantitative definitions of '$<_T$' has been Popper's notions of truth content and falsity content of a theory. In particular he defines the truth content A_T (falsity content A_F) of any theory A as the class of all true (false) sentences which follow logically from A, *i.e.*,

$$A_T = Cn(A) \cap T;\ A_F = Cn(A) \cap F \qquad (2)$$

where $Cn(A)$ is the set of logical consequences of the theory A.[1]

---

[1] In Popper [1963] (p. 233) A_F is defined as in (2), while in Popper [1972] (pp. 49 and 332) A_F is defined differently as a particular relative content. But, as explicitly stated in Popper [1972] (p. 333), the two definitions agree in extension, hence are equivalent. For a proof see proposition *1.4* of Tichý [1974].

Popper leaves it as intuitively obvious what he means by two theories being comparable, or what we will call potentially comparable.[1] Later he is much more explicit and gives a definition (somewhat complicated in form, for methodological reasons)[2] which can easily be shown to be equivalent to the following: two *theories A and B are potentially comparable* (with respect to the truth) if and only if the truth content of A is a subset of the truth content of B or *vice versa*, and the same for falsity contents, *i.e.*,

$$(A_T \subseteq B_T \text{ or } B_T \subseteq A_T) \text{ and } (A_F \subseteq B_F \text{ or } B_F \subseteq A_F).$$

Finally Popper gives essentially the following conditional definition[3]: *if* theories A and B are potentially comparable, then B *corresponds better to the facts* (*or comes closer to the truth*) *than does* A (abbreviated: $A <^1_T B$) if and only if

$$(3) \qquad (B_T \supseteq A_T \text{ and } B_F \nsupseteq A_F) \text{ or } (A_F \supseteq B_F \text{ and } A_T \nsupseteq B_T).$$

Under the assumption that A and B are potentially comparable, clearly (3) is equivalent to

$$(4) \qquad (A_T \subseteq B_T \text{ and } B_F \subset A_F) \text{ or } (A_T \subset B_T \text{ and } A_F \subseteq B_F)$$

(where '$A_T \subset B_T$' means that A_T is a proper subset of B_T, hence $A_T \neq B_T$).

We note that theories A and B are potentially comparable if and only if $Cn(A)$ and $Cn(B)$ are also. And it is just as obvious that

$$A <^1_T B \Leftrightarrow Cn(A) <^1_T Cn(B).$$

Hence in this paper there will be no loss of generality when, for mathematical purposes, we restrict ourselves in the future to theories which are closed under the consequence operation, *i.e.*, theories A such that $Cn(A) = A$.

4 Let us now apply Popper's approach to an arbitrary comparison theory C. As a result we will show that not only does '$<^1_T$' not agree with '$<_T$', but that the difficulty occurs when using any comparison theory C and is not due to T being a limiting case (in some sense) of actual scientific comparison theories.

As in Section 3 we again assume that we are working with theories over a fixed language L. But we make no assumptions about L being a first order language: it could be of higher or lower order. We also do not assume this time that L has a standard interpretation; *e.g.*, L might be the language for some quantum mechanical theories. In keeping with what we said at the end of Section 3, we do restrict a theory to be a consistent set A of statements such that $Cn(A) = A$.

The theories we discuss will be completely arbitrary. Thus we will not assume that the comparison theory C is complete, hence there may be sentences c such that neither c nor $(\neg c)$ is in C. (If L has a truth set T and $C = T$, then of course C is complete.) And we make no assumption about any of the theories being axiomatisable.

For any theory C, let $\overline{C}$ be the set of all sentences whose negations are in C, *i.e.*,

$$\overline{C} = \{c : (\neg c) \in C\}.$$

Intuitively, if C consists of the assumed true sentences, then $\overline{C}$ consists of the assumed-false sentences. Of course in the case when $C = T$ we have $\overline{C} = F$.

---

[1] Popper [1963], p. 233.

[2] Popper [1972], p. 50.

[3] Popper [1963], p. 233.

For any theories A and C, define

$$A_C = A \cap C \text{ and } A_{\bar{C}} = A \cap \bar{C};$$

A_C is the ***assumed-truth content*** and $A_{\bar{C}}$ is the ***assumed-falsity content of theory A with respect to comparison theory C.***

We say that theory A *is false with respect to comparison theory C (or A is C-false)* if and only if

$$A_{\bar{C}} = A \cap \bar{C} \neq \emptyset;$$

and we say that A *is C-true* if and only if $A_{\bar{C}} = \emptyset$. Thus given any theories A and C we will have that A is C-true or C-false.

Let '$a \Leftrightarrow b \Leftrightarrow c$' mean that statements a, b and c are logically equivalent. Then we have the following major result.

Theorem 4.1.[1] For any C-false theories A and B we have

(*a*) $A_C \subseteq B_C \Leftrightarrow A \subseteq B \Leftrightarrow A_{\bar{C}} \subseteq B_{\bar{C}}$

(*b*) $A_C = B_C \Leftrightarrow A = B \Leftrightarrow A_{\bar{C}} = B_{\bar{C}}$

(*c*) $A_C \subset B_C \Leftrightarrow A \subset B \Leftrightarrow A_{\bar{C}} \subset B_{\bar{C}}$.

(Remember: we are assuming that any theory A satisfies $Cn(A) = A$.)

Proof: Clearly (*a*) and (*b*) imply (*c*). Moreover (*a*) implies the non-trivial parts of (*b*) since

$$A_C = B_C \Rightarrow A_C \subseteq B_C \text{ and } B_C \subseteq A_C;$$

and similarly for the case where $\bar{C}$ replaces C. We now prove the non-trivial parts of (*a*).

Claim 1: $A_C \subseteq B_C \Rightarrow A \subseteq B$.

Details: Since B is C-false, there is some element $c \in B \cap \bar{C}$, hence

$$\begin{aligned} a \in A &\Rightarrow (a \vee \neg c) \in A \cap C \subseteq B \cap C \text{ since } (\neg c) \in C \text{ and } A_C \subseteq B_C \\ &\Rightarrow (c \rightarrow a) \in B \cap C \text{ since } (c \rightarrow a) \leftrightarrow (\neg c \vee a) \\ &\Rightarrow a \in B \text{ since } c \in B. \end{aligned}$$

Claim 2: $A_{\bar{C}} \subseteq B_{\bar{C}} \Rightarrow A \subseteq B$.

Details: Since A is C-false, there is some element $c \in A \cap \bar{C}$.

Then for any sentence a we have

$$c \in \bar{C} \Rightarrow (\neg c) \in C \Rightarrow \neg(a \,\&\, c) \in C \Rightarrow (a \,\&\, c) \in \bar{C}$$

hence

$$a \in A \Rightarrow (a \,\&\, c) \in A \cap \bar{C} \subseteq B \cap \bar{C} \Rightarrow a \in B. \qquad \text{Q.E.D.}$$

One striking fact about *Theorem 4.1(c)* is that it says one can't increase the truth content of a false theory by conjoining a logically independent sentence b without also increasing the falsity content and vice versa, whether or not b is true.

---

[1] Equivalent results have been independently discovered by D. Miller (*cf.* his [1974]) for the case $C = T$, hence for the case when C is any complete theory. In Popper [1966], p. 351, there is a proof of $A_T * B_T$ if and only if $A * B$, where $*$ is '$=$', '$\subset$' or '$\subseteq$', for finitely axiomatisable A and B, *whether or not A and B are false*; however, the proof makes essential use of T being a complete non-axiomatisable theory. Miller [1974] has improved Popper's result, requiring only recursive axiomatisability of A and B.

Generalising a definition of Section 3, we say that theories *A and B are potentially comparable with respect to C* if and only if

$$(A_C \subseteq B_C \text{ or } B_C \subseteq A_C) \text{ and } (A_{\bar{C}} \subseteq B_{\bar{C}} \text{ or } B_{\bar{C}} \subseteq A_{\bar{C}}).$$

Then an obvious consequence of *Theorem 4.1(a)* is

Theorem 4.2. Two C-false theories A and B are potentially comparable with respect to comparison theory C if and only if

$$A \subseteq B \text{ or } B \subseteq A.$$

But this result contradicts Popper's and any scientist's idea of comparability. Scientists often compare two theories A and B which are logically incompatible; in such a case we certainly have

$$Cn(A) \nsubseteq Cn(B) \text{ and } Cn(B) \nsubseteq Cn(A).$$

Popper gives an argument for why Newton's theory (N) and Einstein's theory (E) are intuitively (potentially) comparable.[1] But N and E are definitely logically incompatible and are both false, at least with respect to currently accepted experimental values for the bending of light passing the sun (though E seems to be in error by only 3 per cent). Hence according to *Theorem 4.3*, N and E are not formally (potentially) comparable.

Since Popper's definition of potential comparability is faulty, let us not make the generalisation of '$A <^1_T B$' conditional on A and B being potentially comparable. Hence we will just say that *B agrees better with comparison theory C than does A* (abbreviated: $A <^1_C B$) if and only if

$$(A_C \subset B_C \;\&\; B_{\bar{C}} \subseteq A_{\bar{C}}) \text{ or } (A_C \subseteq B_C \;\&\; B_{\bar{C}} \subset A_{\bar{C}}).$$

Then an obvious consequence of *Theorem 4.1(c)* is

Theorem 4.3.[2] If A and B are distinct (*i.e.* $A \neq B$) C-false theories, then

(*a*) $A \not<^1_C B$ and $B \not<^1_C A$;

(*b*) if language L has a truth set T and if $C \subseteq T$, then

$$A \not<^1_T B \text{ and } B \not<^1_T A.$$

Thus if both A and B are C-false, then we don't have $A <^1_C B$ or $B <^1_C A$, even if $A \subset B$. Put another way, of two distinct C-false theories, neither agrees better with the assumed facts of C than the other, if we use the definition of '$<^1_C$'. *Theorem 4.1* tells us exactly why this unacceptable situation arises for Popper's definition.

Let us now consider an interesting example of *theorems 4.3(a)* and *4.1(c)* where the comparison theory is a subset of T.[3] Assume that Newton's theory (N) and

---

[1] Popper [1972], p. 52.

[2] *Theorem 4.3(b)* was first proved independently by D. Miller and P. Tichý, both working in the context of comparison theory T (hence essentially in any complete comparison theory). Tichý's original proof, presented to the philosophy seminar at Otago University in March, 1973, held only for the case when A and B are finitely axiomatisable. In May, 1973 at a meeting of the *Association of Symbolic Logic* I announced a general proof of *Theorem 4.3*. The proof of his corresponding proposition 2.4 in Tichý [1974], though claimed only for finitely axiomatisable A and B, holds as it is without any such restrictions and does trivially generalise to a direct proof of my *Theorem 4.3*.

[3] This example was suggested to me by D. Miller.

Einstein's general theory (E) are both false with respect to the set M of truths of mechanics. (Such would be the case if space-time were found to be quantised.) Then we would have the following consequences:

(*1*) not only are N and E incomparable with respect to T but by *4.3*(*a*) they are already incomparable with respect to M:

$$N \nless_M E \text{ and } E \nless_M N.$$

(*2*) Since $N \nsubseteq E$, we have by *4.1*(*c*) that E does not contain more truths of mechanics than does N, *i.e.*, $N_M \nsubseteq E_M$.

From Popper's writings one sometimes gets the impression that one should exclude logical truths from the truth content.[1] Likewise one might consider having the falsity content contain only contingently false sentences. It can be shown that for any such assumptions, one can still prove a corresponding version of *Theorem 4.1*.[2]

Instead of using both assumed truth content and falsity content one might think of using just one or the other, at least for comparing C-false theories.[3] Say for C-false theories A and B we define

$$A <_C^+ B \Leftrightarrow A_C \subset B_C;\ A <_C^- B^- \Leftrightarrow B_C \subset A^-{}_C.$$

Then of course it follows immediately from *Theorem 4.1*(*c*) that

$$A <_C^+ B \Leftrightarrow A \subset B;\ A <_C^- B \Leftrightarrow B \subset A.$$

Hence no logically incompatible theories could be compared using '$<_C^+$' or '$<_C^-$', a situation discussed in our interpretation of *Theorem 4.2*. But even worse, when '$<_C^+$' and '$<_C^-$' do give comparability, they can give results conflicting with '$<_C$', as the next example shows.

Consider any atomic sentences a and b which are logically independent, *i.e.*, for which $(a \& b)$, $(a \& \neg b)$, $(\neg a \& b)$ and $(\neg a \& \neg b)$ are consistent theories. (In this example we are adopting the following convention:

'$(a_1 \& \ldots \& a_n)$' denotes the theory $Cn(\{a_1 \& \ldots \& a_n\})$.)

Let the comparison theory C be $(a \& b)$. Then for the three C-false theories $(\neg a \& \neg b)$, $(\neg a)$ and $(\neg a \& b)$ we intuitively have

$$(\neg a \& \neg b) <_C (\neg a) <_C (\neg a \& b);$$

yet one easily shows that we have

$$(\neg a \& b) <_C^- (\neg a) <_C^+ (\neg a \& \neg b).$$

JOHN H. HARRIS
University of Otago,
Dunedin, New Zealand

---

[1] *Cf.* Popper [1972], p. 48.

[2] To be specific, if A and B are C-false theories,

$$D \supseteq C - Cn(\phi) \text{ and } E \supseteq \overline{C} - \overline{Cn(\phi)},$$

then we have

$$A \cap D \subseteq B \cap D \Leftrightarrow A \subseteq B \Leftrightarrow A \cap E \cap B \cap E$$

and similarly in the case when we replace '$\subseteq$' by '$=$' or '$\subset$'.

[3] K. R. Popper, after learning of the Miller-Tichý result, *i.e.*, *Theorem 4.3*(*b*), suggested at a seminar at Otago University that one might use just the truth content.

REFERENCES

MILLER, D. [1974]: 'Popper's Qualitative Theory of Verisimilitude' *British Journal for the Philosophy of Science*, **25**, pp. 166–177.

POPPER, K. R. [1963]: *Conjectures and Refutations.*

POPPER, K. R. [1966]: 'A Theorem on Truth-Content' in P. Feyerabend and G. Maxwell (*eds.*): *Mind, Matter and Method, Essays in Philosophy and Science in Honor of Herbert Feigl*, pp. 343–53.

POPPER, K. R. [1972]: *Objective Knowledge.*

TICHÝ, P. [1974]: 'On Popper's Definitions of Verisimilitude', *British Journal for the Philosophy of Science*, **25**, pp. 155–160.

POPPER'S QUALITATIVE THEORY OF VERISIMILITUDE

Popper's theory of verisimilitude was first presented in Popper [1962] and [1963]. It has since been expanded in Popper [1972], especially in Chapters 2 and 9. To say the least, it has been strangely misrepresented and misunderstood by some authors. In Miller [1972], in the course of defending it against one such caricature (Robinson [1971]), I ventured some rather guarded criticism myself. It is the purpose of this paper to deepen and strengthen these objections with regard to the qualitative theory that Popper puts forward; and to show that that theory is quite untenable. Further criticisms that I have of Popper's quantitative theory I leave for another occasion. But since what I have to say will be almost wholly critical, I should perhaps stress that, to my mind, Popper is the only person who has made the slightest progress towards solving the problem of verisimilitude, a problem that he himself more or less discovered. The extent of my criticisms would be without point if I did not think that there was here a difficult problem to be solved. I should be happy myself to provide a theory of nearness to the truth a little nearer to the truth than his own theory of nearness to the truth.

1 Few of the scientific theories that we discover are true; perhaps none is. All anyway appear to be oversimplifications. Yet theories of motion, to take a familiar example, starting with Plato and Aristotle and developing through Buridan and Oresme, Galileo and Descartes, to Newton and Einstein, do give the appearance (deceptive perhaps) of progress—of progress towards the truth. What can there be about one false theory that makes it closer to the truth than is another?

This is the problem of verisimilitude.

Popper's intention was to show that 'the idea of nearness to the truth, or of a better or less good approximation to the truth [,] . . . the idea of "*verisimilitude*" [,] . . . can be rehabilitated in a way similar to Tarski's rehabilitation of the idea of truth as correspondence to the facts' ([1970], p. 29; [1972], p. 318). He stressed that the 'idea of approximation to truth, or of verisimilitude, has the same objective character and the same ideal or regulative character as the idea of objective or absolute *truth*. It is *not an epistemological or an epistemic idea*—no more than [is] truth or content' ([1963], p. 234). Thus verisimilitude (unlike degree of corroboration, say, or even logical probability, or credence) is not a function of the evidence, or of what evidence is available; though like truth itself it may be a function of the truth value of the evidence. This does not mean that we may not, taking account of the evidence, hazard a guess at how close a theory is to the truth; but it does mean that such a guess, like a guess at a theory's truth value, is, as Hume so accurately put it, 'entirely precarious' ([1777], p. 27). And the same goes, or can be expected to go, for conjectures concerning the relative proximity of two theories to the truth.[1]

What can there be about one false theory that makes it closer to the truth than is another? A crude reply might be: the first theory is logically stronger than the second. Newton's theory makes progress over Galileo's simply because it says more than Galileo's theory says.

Now there can be no doubt that the content, or strength, or power of a deductive theory is determined by the size of its consequence class; the more consequences a theory has, the stronger it is. Thus one way in which a theory can grow towards encompassing the whole truth is simply through an increase in content. But, it can obviously be objected, and rightly, mere aggregation of consequences is of no use in itself; we want these new consequences to be *true*. After all, any old fairy tale, however preposterous, if grafted on to Galileo's theory, will beget new consequences. But we would be disinclined to accept such an augmented theory, in contrast to Newton's, as a step in the direction of the truth.[2] Or to put the matter in another way: increase in truthlikeness surely depends to at least some extent on what is true. But the suggestion would make it quite independent of the truth.

This elementary consideration led Popper to separate the true consequences of a theory—its *truth content*—from the false consequences—its *falsity content*—, and to suggest that a theory B be regarded as closer to the truth than is a theory A if it either surpasses it in truth content, without also surpassing it in falsity content, or it lags behind A in falsity content, but not in truth content. In Popper [1963], p. 233, he says:

> *Assuming that the truth-content and the falsity-content of two theories t_1 and t_2 are comparable, we can say that t_2 is more closely similar to the truth, or corresponds better to the facts, than t_1, if and only if either*
> (*a*) *the truth-content but not the falsity-content of t_2 exceeds that of t_1,*
> (*b*) *the falsity-content of t_1, but not its truth-content, exceeds that of t_2.*

For brevity I shall call this proposal (◇). A similar proposal appears in Popper [1972], p. 52.

This theory is purely qualitative, for only if one truth content (or falsity content) properly includes another (in the usual set-theoretical sense) can we say that it exceeds that other. So the theory is obviously rather restricted too: it would come as no surprise if many pairs of theories we should like to compare for verisimilitude turned out in fact not to be comparable either by truth content or by falsity content. Popper was of course aware of this restrictiveness, and his quantitative theory was introduced partly in order to overcome it. But there is not much doubt that the passage quoted above was meant to express the intuitive background of the whole theory of verisimilitude, quantitative as well as qualitative. That is to say, the method of comparing truth contents and falsity contents was meant to equip us with simple, though perhaps not very realistic, examples of how false theories could be compared for verisimilitude. But in fact the method provides no such examples. It will be shown in 3 below that *no two false theories are comparable by verisimilitude* in this way; indeed, if we require axiomatisability, *only true theories are so comparable.* And neither can we really say that these considerations provide us with 'the logical basis of the method of science—the method of bold conjectures and of attempted refutations' (*op. cit.*, p. 53). Having shown that if B exceeds A in content then it exceeds it also in truth content,[1] so that 'the theory with the greater content . . . will also be the one with the greater verisimilitude *unless its falsity content is also greater*' (*loc. cit.*), Popper thought that he had explained the desirability of seeking falsifications. Everything, it seems, depends on the falsity content (I thought this too when he first discussed it with me). But, indeed, if B exceeds A in content then either it exceeds it also in falsity content or they are both true. Thus the searching out of falsity, if a good method (as I think it is), is not to be vindicated in this manner.

---

[1] In this I seem on the face of it to disagree with Watkins, who writes that there can be '*good reasons for claiming* that a particular hypothesis . . . is closer to the truth than a rival hypothesis' ([1968], p. 280).

[2] This is not to say that an increase in content is not, methodologically speaking, always an improvement (at least potentially). We are talking of logic, not method.

My objection to this qualitative theory is therefore not that few false theories are comparable by its canons but that no two are. That these are independent points can be seen by advancing to a quantitative version of the theory. Popper shows how to measure truth content and falsity content so that any two theories are comparable by both. Thus all problems of *comparability in principle* are overcome. Yet if we compare B and A separately by truth content and falsity content, and ask for the quantitative version of condition ($\Diamond$) to hold, we obtain the same result as before: B must be true. Thus *no two false theories are comparable* by verisimilitude *in this way*.[2]

These results, along with some others, are all contained in 3–5. In 2 I present the logical essentials needed for the understanding of these sections.

2 At a pinch we can think of all scientific theories as formulated in some first-order language or other. If so formulated they cannot be expected to be finitely axiomatisable (provided they are consistent); this is one of the things that Gödel's theorem tells us. Consequently our discussion is best conducted against the backcloth of Tarski's *calculus of deductive systems*, a deductive system being any set of sentences closed under the operation of logical consequence. Following Tarski I shall use Cn for the consequence operation. In this notation a deductive system is a set of sentences A which satisfies $A = Cn(A)$. Systems can be partially ordered by the usual relation of set-theoretic inclusion; and under this ordering the set S of all sentences becomes the greatest system of all, whilst L, the set of logical truths, is the smallest. The intersection $A \cdot B$ of any two systems A and B is a system (as is the intersection of any family of systems); and for any two systems A and B there is a smallest system $A + B$ which includes them both. Likewise, if A is a system there is a contradictory system $-A$, the largest system B for which $A \cdot B = L$.

---

[1] This is not quite Popper's result of [1966]. I have simplified it so as not to bog down the argument. For a generalisation see COROLLARY 2 to THEOREM 5 in 5 below.

[2] I say '*in this way*' because of course Popper gives a quantitative definition of verisimilitude in terms of truth content and falsity content, and according to this a false B can have a greater verisimilitude than a suitable A. But this is quite a different matter.

Thus deductive systems form a lattice. Indeed, the $\subseteq$-ordering on the class of deductive systems imposes on it the structure of a Heyting algebra with zero L and unit S. (See Tarski [1935], section 2, for this result. Heyting algebras are called pseudo-Boolean algebras in Rasiowa and Sikorski [1963], Chapter 1, section 12.) A Heyting algebra which is not a Boolean algebra bears the same relationship to a Boolean algebra that intuitionist logic bears to classical. For example, it is not in general true that $A + -A = S$ (the 'law of excluded middle'), nor that $- -A = A$. The lattice of finitely axiomatisable systems (sentences), however, is well known to be a Boolean algebra, so the discrepancy in the general case is solely attributable to the presence (or presumptive presence) of non-axiomatisable systems. To see why, consider the language of Peano arithmetic. Gödel's theorem says that the system P (Peano's arithmetic) is, if consistent, neither complete nor completable: no consistent recursively axiomatisable extension of P contains all true sentences of arithmetic. So T, the system of all these true sentences, is not axiomatisable (even recursively). Now what is $-T$? By stipulation $-T$ is the largest system sharing with T only logical truths amongst its consequences. It is not difficult to see that this is L itself.[1] But then $T + -T$, the smallest common extension of T and $-T$, is T, not S; and $---T$ is S, not T.

A system which for any sentence a contains either a or $\urcorner a$ (but not both) is called *complete*. It is important and obvious that T, the set of all true sentences, is complete. There are few interesting cases where T is even recursively axiomatisable, in view of Gödel's theorem. But axiomatisable complete theories exist in many languages (see Tarski [1936], section 4, p. 370). If a is one such, its complement $\urcorner a$ is called by Tarski *irreducible*. An irreducible system contains no proper subsystem but L.

Henceforth I shall normally use 'axiomatisable' in the sense of 'recursively axiomatisable'. If I want to indicate that a system is finitely axiomatisable I shall call it 'finitely axiomatisable', or refer to it as a 'sentence'. This should not be taken to mean that I do not, where necessary, appreciate the difference between a sentence and the class of its consequences.

It is to be noted that the product $A \,.\, B$ of two systems is logically not stronger than either—logically speaking it is their *sum*; likewise the sum $A + B$ is at least as strong as both A and B, and has the force of a logical *product*. In other words, the lattice ordering is an ordering by logical strength, rather than, as in the case of sentential logic, by logical weakness. Since I am a lot more interested in this paper in the logical power of deductive systems than I am in their algebraic or set-theoretical interrelations, I shall from here on drop the algebraic notation

---

[1] This is a corollary to THEOREM 4 in section 5. But it amounts to no more than this. Suppose that there is some untautological consequence a in $-T$. Then a itself can share no contingent consequence with T. Thus for every such consequence b of T ($0 \neq 1$, for example, or Lagrange's theorem) $a \vee b$ must be a logical truth. By sentence calculus, $\urcorner a \vdash b$. So $\urcorner a$ is a consistent finitely axiomatisable system as strong as T, and therefore a finite axiomatisation of T, which is not possible.

Incidentally, since it follows easily from a result of Montague [1956] that no consistent finitely axiomatisable theory is stronger than Zermelo/Fraenkel set theory (*ZF*), the above proof shows that $-ZF$ is also L. (The example of *ZF* moreover settles in the negative an open problem of Popper [1966], p. 352.)

($+$, ., $-$, $<$, $=$) in favour of sentential connectives (&, $\vee$, $\neg$) and the symbols $\vdash$ (unilateral derivability) and $//$ (interderivability). I shall also coin, by analogy with $\leqslant$, the symbol $\Vdash$ for the weak derivability relation (that is, $\vdash$ *or* $//$). These will all be used for sentences and non-axiomatisable systems indiscriminately. Thus $A \vee B$ is the product of the systems A and B, and contains all the consequences common to them. It therefore holds just in case one or other of A and B holds. In the same way, the system $T + -T$, transliterated into this notation, becomes $T \mathbin{\&} \neg T$.

In general, lower-case romans will be used for sentences. Upper-case romans will be variables for sets of sentences, though unless otherwise stated it will be assumed that these sets constitute systems. Letters at the end of the alphabet will be reserved for complete systems. I shall not assume that I am dealing with any particular first-order language, but will use the constants L (system of logical truths) and T (system of all true sentences) in a systematically ambiguous manner.

The truth content of a system A is the intersection of A with T, is therefore $A \vee T$. It consists of all the true consequences of A. The falsity content of A —the class of all its false consequences—is of course not a deductive system: it contains no tautologies. I follow Popper in writing $Ct_F(A)$ for the falsity content of A. (Analogously, I shall sometimes write $Ct_T(A)$ for the truth content of A.)

Two algebraic structures are said to be elementarily equivalent if the sentences that hold in the one hold in the other. By the upward Löwenheim/Skolem theorem a system that holds in any infinite structure at all holds in structures of every infinite cardinality. Thus the class of structures in which a system A holds is not a set. The most reasonable way of reducing it to a set is to identify structures that are elementarily equivalent. Each structure will then correspond to a complete theory, and there cannot be more than continuum many complete theories in a countable language. If A holds in a structure, we will say that that structure is a *model* of A. We will not talk of sentences' or systems' being 'true' in arbitrary structures, for we will want to identify one particular structure ('the world') as the unique one in which T, the true system, holds. If we write $\mathscr{M}(A)$ for the set of all models of the deductive system A it is easily seen that $\mathscr{M}(A \vee B) = \mathscr{M}(A) \cup \mathscr{M}(B)$, and that $\mathscr{M}(\neg a) = -\mathscr{M}(a)$. But the latter result does not hold for unaxiomatisable systems in general: $\mathscr{M}(T)$, for example, is a singleton, but since $\neg T$ is L, $\mathscr{M}(\neg T)$ contains all structures.

3 In this section we uncover just what the circumstances are in which set-theoretical comparisons of truth content and falsity content are feasible; and in particular when there exist two distinct systems A and B comparable by ($\Diamond$) in terms of truthlikeness. When, that is, we have

$$(\Diamond) \qquad \begin{array}{ll} \textit{either} & Ct_T(A) \subset Ct_T(B) \textit{ and } Ct_F(B) \subseteq Ct_F(A) \\ \textit{or} & Ct_T(A) \subseteq Ct_T(B) \textit{ and } Ct_F(B) \subset Ct_F(A). \end{array}$$

The conclusion is:[1] B must be a true theory and A a theory (true or false) whose truth content $A \vee T$ follows from B. Furthermore, if T is not axiomatisable, but B is, then A must also be true.

---

[1] This result dates from July 1972 (THEOREM 5 in its present form was not proved generally until July 1973, though its COROLLARY 2 was established, in a less lucid manner, in 1967). Most of the results of sections 3–5 were announced at the meeting of

For B to be closer to the truth than A is it must either have more truth content or less falsity content. Yet both these seem, at a casual inspection, to vary somewhat directly with content. Popper [1966] anyway proved that in normal circumstances truth content increases with content. And as for falsity content —suppose that $b \vdash a$ and b is false; in this case all a's falsity content is in b's, and b is besides. Thus b outstrips a in falsity content.

In THEOREMS 1–4 we investigate exactly how truth content and falsity content vary with content. We first consider theories comparable by falsity content.

THEOREM 1: Let A be false. If $Ct_F(A) \subseteq Ct_F(B)$ then $B \Vdash A$.

Proof. Suppose that c is a false consequence of A (there is such a c, for A is false) and that a is any consequence of A. Then a & c is in $Ct_F(A)$, so in $Ct_F(B)$; therefore a & c is a consequence of B, which means that a is, as required.

COROLLARY 1: If A is false and $Ct_F(A) \subset Ct_F(B)$ then $B \vdash A$.

Proof. Were B equivalent to A it would have the same falsity content.

COROLLARY 2: If A is false and $Ct_F(A) = Ct_F(B)$ then $B \mathbin{/\!/} A$.

Proof. The assumptions suffice to guarantee that B is also false. Thus we can apply the theorem in each direction.

Thus two false theories comparable by falsity content are also comparable by content. (The assumption of falsity is obviously needed.) There is something like a converse to this theorem.

THEOREM 2: If $B \mathbin{/\!/} A$, or $B \vdash A$ and B is true, then $Ct_F(B) = Ct_F(A)$.
If $B \vdash A$ and B is false, then $Ct_F(A) \subset Ct_F(B)$.

Proof. The first assertion is trivial. For the second, we certainly have $Ct_F(A) \subseteq Ct_F(B)$. Equality would, however, by COROLLARY 2 to THEOREM 1, mean that B and A were interderivable.

We now turn to the problem of how content varies with truth content. We note that $B \vee T \Vdash A \vee T$ is just another way of saying that $A \vee T \subseteq B \vee T$, that A's truth content is included in B's.

THEOREM 3: Let B be false or A be true. If $B \vee T \Vdash A \vee T$, then $B \Vdash A$.

Proof. Since $B \vee T \Vdash A \vee T$ we must have $B \Vdash A \vee T$. Now suppose that not $B \Vdash A$. Accordingly A must fail in some model of B, and so T must hold

---

the Association for Symbolic Logic in Bristol in July 1973. (More rudimentary versions of the paper were read at seminars at the London School of Economics in February 1973 and at the University of Warwick in June 1973.) That when A and B are finitely axiomatisable B must be true was already known to Professor C. G. Hempel in 1970 (so Professor J. Hattiangadi has recently informed me). The more general result has also been discovered quite independently by Dr P. Tichý, whose paper (Tichý [1974]) I received on my return from Bristol. Tichý's proof of the result (his *Proposition* 2.4) is particularly elegant. A generalisation of Tichý's *Proposition* 2.4 has been given by Dr J. Harris (Harris [1974]).

in that model. But T has only one model. Thus B holds and A fails in T's only model; so B is true and A is false.

Consequently, if either B is false or A is true, we must have $B \Vdash A$.

COROLLARY 1: If B is false or A is true and $B \vee T \vdash A \vee T$, then $B \vdash A$.

Proof. Were B and A interderivable $B \vee T$ would be interderivable with $A \vee T$.

COROLLARY 2: If A and B have the same truth value and $A \vee T \mathbin{/\!/} B \vee T$, then $A \mathbin{/\!/} B$.

Proof. Material equivalence of A and B means that 'B is false or A is true' and 'A is false or B is true' both hold. Thus the THEOREM can be applied to each in turn.

COROLLARY 3: If $B \vee T \Vdash A \vee T$, then $Ct_F(A) \subseteq Ct_F(B)$ or $Ct_F(B) \subset Ct_F(A)$.

Proof. This is a simple combination of THEOREMS 2 and 3.

These results are enough to show the inadequacy of condition ($\Diamond$), as follows. If A and B are to be comparable by verisimilitude they must be (minimum condition) comparable by truth content. According to the last corollary they will then be comparable by falsity content. For interest's sake we suppose that A and B are not identical, and for definiteness's that $B \vee T \Vdash A \vee T$; that B does not fall short of A in truth content. Then one of the four following possibilities must obtain:

(*a*) B is true and A is true. Thus $B \Vdash A$, and so $B \vdash A$ (since they are not identical). Thus $B \vee T \vdash A \vee T$. We conclude that B is closer to the truth than A is.

(*b*) B is true and A is false. A definitely has more falsity content than does B. Thus B is closer to the truth than A is.

(*c*) B is false and A is false. We can apply THEOREM 3 and conclude that $B \Vdash A$; so $B \vdash A$. THEOREM 2 now tells us that B exceeds A in falsity content. Since A is false, we can use COROLLARY 2 to the last theorem: it follows that B does not have the same truth content as does A, so has more. Thus B exceeds A in both truth and falsity content, so that they are not comparable.

(*d*) B is false and A is true. As in (*c*), B exceeds A in falsity content. If their truth contents are not the same, then B has more, so that no comparison is possible. Otherwise we have a subcase of (*b*) with A and B reversed.

It can now be seen that all the cases in which verisimilitudes are comparable fall under the rubric already given: for B to be nearer the truth than A is, B must be a true theory and $A \vee T$ must follow from it.

4 Might it be that this result merely reflects the widespread incomparability of truth contents and falsity contents? After all, set-theoretical inclusion is a relation that obtains somewhat seldom, and it is therefore less than surprising that few pairs of theories are comparable by verisimilitude. I intend to show briefly that this is not the case. For we can adopt quantitative measures of truth content and falsity content that allow any two contents to be compared. Yet it turns out that in simple finite cases for B to be closer to the truth than A is, once again B must be true.

The measures I shall use are those suggested and explained by Popper himself ([1963], pp. 393–6; [1972], pp. 51 and 334), and are based on his 1954 measure of content in terms of logical probability. The content of A is equal to the ***improbability*** of A, that is $1 - p(A)$. The truth content of A, reasonably enough, is $1 - p(A \vee T)$, and the falsity content is $1 - p(A, A \vee T)$. For convenience I use $\alpha, \beta, \ldots, \tau, \ldots$ for the probabilities of the systems $A, B, \ldots, T, \ldots$. If A is true, its truth content then has measure $1 - \alpha$, its falsity content has measure 0. If A is false it of course contradicts T, so that $p(A \vee T) = p(A) + p(T) = \alpha + \tau$. Thus the truth content of a false A is $1 - (\alpha + \tau)$. We can also compute the falsity content if α and τ are not both zero: the value is $1 - \alpha/(\alpha + \tau) = \tau/(\alpha + \tau)$.[1]

Now suppose that B is false. If $\tau \neq 0$ condition ($\Diamond$) certainly cannot hold between B and any true A. And if A is false, it reduces to: ***either*** $\beta + \tau < \alpha + \tau$ and $\tau/(\beta + \tau) \leqslant \tau/(\alpha + \tau)$ ***or*** $\beta + \tau \leqslant \alpha + \tau$ and $\tau/(\beta + \tau) < \tau/(\alpha + \tau)$.

Each of these is impossible. Thus even in the case of a finite language, where Popper's measures of truth content and falsity content are totally reasonable, no two false theories can be persuaded to be compared for verisimilitude by ($\Diamond$). It seems clear that it is ($\Diamond$) that is to blame.

5 Two tasks remain from section 3. One is to provide something of a converse to THEOREM 3—to show how increase in truth content depends on increase in content. The other is to provide an illustration of our conclusion—a true system B from which we can derive the truth content of a false system A (there is no difficulty in finding a true system A).

THEOREM 4 will provide the converse to THEOREM 3. A lemma (which is in fact equivalent to THEOREM 4) will be generalised to THEOREM 5, which will show that no concrete or real-life illustration is possible: that if A is false then $A \vee T$ has no true extension B that is axiomatisable.

These results all depend on the unaxiomatisability of T. This will normally be the case in any language that includes arithmetic. But it may easily be the case for other reasons too.

LEMMA: Let T be unaxiomatisable. If A is false then $A \vee T$ is unaxiomatisable.

Proof. Let a be some false consequence of A. Were $A \vee T$ axiomatisable, so would $(A \vee T)$ & $\neg a$ be. But we can easily prove this theory to be equivalent to T.

It certainly is true, so included in T. Moreover, if c is true so is $a \vee c$, which therefore belongs to $A \vee T$. By sentential logic $A \vee T, \neg a \Vdash c$. Thus $A \vee T$, together with the sentence $\neg a$, entails every true sentence.

THEOREM 4: Let T be unaxiomatisable. If A is axiomatisable and $B \vdash A$, then $B \vee T \vdash A \vee T$.

---

[1] If $\tau = 0$ other considerations can be brought to bear (see, for example, my [1972], pp. 54 f.). We should notice that a zero value for τ allows false theories to have the same falsity content as true theories. It is therefore to be doubted if $1 - p(A, A \vee T)$ is an entirely adequate measure of falsity content in the infinitistic case.

The simple calculation in the text (which has also been performed by P. Tichý) shows that, contrary to Hesse [1968], p. 332, something can be said about how measures of falsity content vary with measures of content. On the other hand, the results of 3 show her criticisms not to have been misplaced (even if the illustration she uses is somewhat unconvincing).

Proof. Certainly $B \vee T \Vdash A \vee T$. If the truth contents were interderivable we would have $B \vdash A \Vdash B \vee T$. Now B and $B \vee T$ differ by at most one model, and the number of models must *increase* across $\vdash$. Thus $A \,//\, B \vee T$. By the LEMMA B is true. So $A \,//\, B$, which contradicts the assumption that $B \vdash A$.

COROLLARY: Let T be unaxiomatisable. The system contradictory to it is the system L. (See also *above*, p. 169, footnote 1.)

Proof. If $\neg T$ is not the same as L, then $\neg T \vdash L$. L is axiomatisable. Thus $\neg T \vee T \vdash L \vee T$. But $\neg T \vee T \,//\, L$ by the definition of the contradictory. And it is not the case that $L \vdash L \vee T$. Consequently $\neg T \,//\, L$.

What the LEMMA tells us is that if A is false there is no (true) axiomatisable B such that $B \,//\, A \vee T$. Thus if an axiomatisable B is to be closer to the truth than is a false A it will have strictly to entail $A \vee T$. We now show that this is not possible either. That is, if B is axiomatisable (not a particularly punishing requirement for a scientific theory) and true, then $B \vdash A \vee T$ only if A is also true. The proof in fact is a trivial extension of the proof of the LEMMA.

THEOREM 5: Let A be false and B be true. If $B \Vdash A \vee T$ then $B \;\&\; \neg a \,//\, T$ for every false a such that $A \Vdash a$.

Proof. $B \;\&\; \neg a$ is true. Moreover, if c is true than $a \vee c$ is derivable from $A \vee T$, and so from B. Thus c is derivable from $B \;\&\; \neg a$.

COROLLARY 1: Let T be unaxiomatisable, and B be axiomatisable. If $B \vdash A \vee T$ and B is true, then A is true.

Proof. If A is false then $B \;\&\; \neg a$ is unaxiomatisable, so that B must be too.

COROLLARY 2: Let T be unaxiomatisable, and A and B axiomatisable. Then $B \vdash A$ if and only if $B \vee T \vdash A \vee \mathrm{T}$, and $B \,//\, A$ if and only if $B \vee T \,//\, A \vee T$.

Proof. THEOREM 4 and the laws of identity deliver the left-to-right implications. Since B is axiomatisable, if it is true, so is A, by the previous COROLLARY. The right-to-left implication is then trivial. If B is false, COROLLARY 1 to THEOREM 3 assures us that if $B \vee T \vdash A \vee \mathrm{T}$ then $B \vdash A$. And if A is also false, COROLLARY 2 to that THEOREM assures us that if $A \vee T \,//\, B \vee T$ then $A \,//\, B$. The only remaining case is of B false and A true, and $A \vee T$ being interderivable with $B \vee T$. Since A is axiomatisable this is ruled out by the LEMMA.

A special case of this last COROLLARY, where A and B are both assumed to be finitely axiomatisable was first proved by Popper in his [1966]. It is worth noting (as he has noted) that the LEMMA follows from THEOREM 4; it suffices to write $B \vee T$ for A.

The upshot of COROLLARY 1 to THEOREM 5 is that no two distinct axiomatisable theories may be compared for verisimilitude unless they are both true. But in this case we shall hardly need the sophistications of verisimilitude to tell us which of the systems we would in principle prefer. For one of them will be logically stronger than the other, so clearly an improvement.

I conclude that the qualitative theory of truth contents and falsity contents is an inadequate theory; and I doubt whether it can serve, even intuitively, as the basis of a richer theory. To substantiate this would, however, require another paper.

6 The above sections were more or less in their present form before I received the papers of Tichý [1974] and Harris [1974] which also appear in this number of the *Journal*. I am grateful to have been able to see these papers, and to correspond with both Tichý and Harris. With much of what they say I of course concur. But there is one serious point on which I disagree with both of them. In consequence, I should like to append here a few critical remarks.

Both Tichý and Harris claim to be able to make, in certain very simple situations, correct judgments of the comparative proximity to the truth of different theories. (See section 3 of Tichý's paper, and the last paragraph of Harris's.) To my mind, however, *the plausibility of these judgments is entirely misleading.* If I am right in this, then there will in fact be no very damaging consequences for Harris's position. For he only uses these assessments to argue that *having more truth content* and *having less falsity content* cannot act as adequate substitutes for ($\Diamond$). (His point is stated more generally, in terms of 'comparison theory' C, of which T would be a special case. But this need not concern us here.) He asserts, that is, that each of the weakened conditions suggested clashes with intuitively obvious assessments of truthlikeness. This fact, however, can easily be established without recourse to these assessments, which are neither intuitively obvious nor even correct. For were B's having less falsity content than A a sufficient condition for it to be closer to the truth, it would follow that every true theory is closer to the truth than is every false theory; so that a tautology would be more truthlike than Newton's theory. On the other hand, if additional *truth content* were a sufficient condition for greater verisimilitude, then amongst axiomatisable theories additional *content* would be enough (by THEOREM 5, COROLLARY 2). And this possibility we have already rejected; see the text to footnote 2, p. 167 *above*.

Much of Tichý's paper, however, will be adversely affected, including the second part of his criticism of Popper's quantitative theory of verisimilitude. (With the earlier part of his criticism I largely agree.) Also undermined is the sketch of a new theory that Tichý offers in the final two paragraphs of his paper. I shall for this reason trim my criticism to these parts of Tichý's contribution, and ignore its (very straightforward) adaptation to Harris's paper. In the course of doing this I shall make some minor adjustments to Tichý's terminology, largely to make the exposition more graphic, and thereby more clear.

Tichý considers (p. 158) 'a rudimentary weather-language L containing no predicates and only three primitive sentences'. Let these sentences be 'It is hot', 'It is raining', and 'It is windy', abbreviated respectively by h, r, w. Tichý supposes all three sentences to be true, so that the whole truth, so far as this tiny language can grasp it, is $h \,\&\, r \,\&\, w$. He shows, though his proof is independent of any particular example, and, indeed, of any particular (strictly positive) probability measure, that according to Popper's quantitative theory every false theory with the same content has the same falsity content, and also the same verisimilitude. (See the second paragraph of my section 4.) So if we take a uniform measure over the maximal consistent sentences, that is, the constituents, every constituent apart from the true one ($h \,\&\, r \,\&\, w$) will have the same verisimilitude.

Tichý goes on to imagine (p. 159) that 'Jones and Smith, two prisoners sharing a windowless and air-conditioned cell, are using L to discuss the weather. Jones takes the view that it is a dry, still day, with a low temperature'; in other words, Jones's conjecture is $\neg h \& \neg r \& \neg w$. 'Smith disagrees. Although he also thinks that the temperature is low, he (rightly) insists that it is raining and windy.' In other words, Smith's theory is $\neg h \& r \& w$. Tichý now suggests that since Jones is 'wrong on three counts', whilst Smith is wrong on only one, Smith's theory is closer to the truth than Jones's. Yet Popper's measure of verisimilitude gives the same value for each theory, on the basis, as I say, of a uniform measure over the constituents. (Tichý does not consider any other possibility.)

But is Tichý right to suppose it 'hardly deniable' that Smith is nearer to the truth than Jones is? I do not really think so. Consider another rudimentary weather language, trivially intertranslatable with L. This language also has three atomic sentences. One is h ('It is hot'). Another is m ('It is Minnesotan'); the third is a ('It is Arizonan'). The weather in Minnesota, we may suppose, is either hot and wet or cold and dry. Thus m is equivalent to $h \leftrightarrow r$. Arizonan weather, on the other hand, is either hot and windy or cold and still. Thus a is equivalent to $h \leftrightarrow w$. The three sentences h, m, a are logically independent of one another; and the eight constituents that can be formed from the three generators h, m, a are just the eight constituents that can be formed from h, r, w. Moreover, r is equivalent to $h \leftrightarrow m$, and w is equivalent to $h \leftrightarrow a$. There is perfect symmetry. And there seems no good reason—beyond sheer prejudice—for treating the *h–r–w* language as more fundamental than the *h–m–a* one. Yet Smith's theory in the latter language is $\neg h \& \neg m \& \neg a$, as is easily checked, whilst Jones's is $\neg h \& m \& a$. The truth comes out as $h \& m \& a$. Thus Smith is wrong on three counts, Jones on only one. Which one now is nearer to the truth?

What does seem to me 'hardly deniable' is that the theories $h \& m \& a$ and $h \& r \& w$ are equivalent theories (and, likewise, each of the seven other constituents in the *h–m–a* language is equivalent to a constituent in the *h–r–w* language). They have, after all, a common definitional extension (see Kanger [1968]). This might indeed incline us to say that they are not so much equivalent as identical. If I understand Hanson correctly, this would have been his view (see his [1961], p. 403). But whatever we do say, we can hardly allow comparisons of verisimilitude to be as language-dependent as they are under Tichý's 'counting method'. Verisimilitude, like truth, will always, if defined, be defined relative to a language. But just as truth is language-independent (this is one of the things that Tarski's T-schema insists on), so must judgments of verisimilitude be, if they are to have any objective significance at all.[1]

This reaction to Tichý's paper does not of course in any way show that Popper's quantitative theory is satisfactory. Nor is it meant to. But it does explain why I cannot accept Tichý's own definition of verisimilitude as the average distance from the true constituent of the constituents in the disjunctive normal form of a sentence. And his projected elaboration of this idea in terms of Hintikka's distributive normal forms is more than likely, I think, to fall foul of the same sort of objection. It is a well-known failing of the confirmation theories of Carnap and Hintikka that degrees of confirmation for sentences are in almost all cases language-dependent.[1] To be sure, this language-dependence is usually demonstrated by splitting some primitive predicate into two independent parts; in which case, the perfect symmetry of the above intertranslations is lacking. But the language-dependence is there all right, and can only be removed, more or less, by the accumulation of vast quantities of evidence. Verisimilitude, however, is not evidence-dependent in this way. It is to be hoped, therefore, that theories of verisimilitude can from the start be free from this particular defect.

---

[1] This paragraph has been inserted to answer some points raised by Dr Tichý in correspondence.

DAVID MILLER
University of Warwick

REFERENCES

HANSON, N. R. [1961]: 'Are Wave Mechanics and Matrix Mechanics Equivalent Theories?', in H. Feigl and G. Maxwell (*eds.*): *Current Issues in the Philosophy of Science.*

HARRIS, J. H. [1974]: 'Popper's Definitions of "Verisimilitude" ', *British Journal for the Philosophy of Science*, **25**, pp. 160–66.

HESSE, M. [1968]: Review of parts of P. K. Feyerabend, and G. Maxwell, (*eds.*): *Mind, Matter and Method*, in *British Journal for the Philosophy of Science*, **18**, pp. 330–5.

HUME, D. [1777]: *An Enquiry Concerning Human Understanding.* References are to the second edition of L. A. Selby-Bigge.

KANGER, S. [1968]: 'Equivalent Theories', *Theoria*, **34**, pp. 1–6.

MILLER, D. [1972]: 'The Truth-likeness of Truthlikeness', *Analysis*, **33**, pp. 50–5.

MONTAGUE, R. [1956]: 'Zermelo-Fraenkel Set Theory is not a Finite Extension of Zermelo Set Theory', *Bulletin of the American Mathematical Society*, **62**, p. 260.

POPPER, K. R. [1962]: 'Some Comments on Truth and the Growth of Knowledge', in E. Nagel, P. Suppes, and A. Tarski (*eds.*): *Logic, Methodology and Philosophy of Science*, pp. 285–92.

POPPER, K. R. [1963]: *Conjectures and Refutations.*

POPPER, K. R. [1966]: 'A Theorem on Truth-Content', in P. K. Feyerabend and G. Maxwell (*eds.*): *Mind, Matter and Method*, pp. 343–53.

POPPER, K. R. [1970]: 'A Realist View of Logic, Physics, and History', in W. Yourgrau and A. D. Breck (*eds.*): *Physics, Logic and History*, pp. 1–30.

POPPER, K. R. [1972]: *Objective Knowledge.*

RASIOWA, H. and SIKORSKI, R. [1963]: *The Mathematics of Metamathematics.*

ROBINSON, G. S. [1971]: 'Popper's Verisimilitude', *Analysis*, **31**, pp. 194–6.

TARSKI, A. [1935]: 'Grundzüge des Systemenkalkül, Erster Teil', *Fundamenta Mathematicae*, **25**, pp. 503–26. References are to the translation by J. H. Woodger in Tarski [1956].

TARSKI, A. [1936]: 'Grundzüge des Systemenkalkül, Zweiter Teil', *Fundamenta Mathematicae*, **26**, pp. 283–301. References are to the translation by J. H. Woodger in Tarski [1956].

TARSKI, A. [1956]: *Logic, Semantics, Metamathematics.*

TICHÝ, P. [1974]: 'On Popper's Definitions of Verisimilitude', *British Journal for the Philosophy of Science*, **25**, pp. 155–60.

WATKINS, J. W. N. [1968]: 'Hume, Carnap and Popper', in I. Lakatos (*ed.*): *The Problem of Inductive Logic*, pp. 271–82.

---

[1] It is also, of course, a failing of the principle of indifference in its classical form.

ON THE COMPARISON OF FALSE THEORIES BY THEIR BASES

My problem is a straight problem in logic; but much of its interest accrues, I believe, from the negative impact it has on an ingenious recent suggestion in the theory of *verisimilitude* or *truthlikeness*. Accordingly, at the outset I shall merely state the logical problem and the bones of its solution, proceeding at once to give an outline of the philosophical setting in which it has been nurtured (section 1).

The problem is concerned with independent sets of sentences—what Tarski[1] has called *bases*. Suppose that C is a set of true sentences, D is a non-empty set of false sentences, and that $C \cup D$ provides a basis (an 'independent axiomatisation') of the *false* theory B. *Under what conditions can some of the sentences in* $C \cup D$ *be exchanged for sentences of the opposite truth value, without sacrifice of either logical strength or independence?*

More formally, let C_0 be a subset of C; (1) *in what circumstances is there a basis for* B *whose true elements are those of* C_0 *and whose false elements include all those of* D? Likewise, let D_0 be a subset of D; (2) *in what circumstances is there a basis for* B *whose false elements are those of* D_0 *and whose true elements include all those of* C?

The answer to question (2) is that there will be such a basis if and only if the set D_0 is not empty (the necessity of this condition is immediate). Question (1) is more difficult. A sufficient condition for the existence of such a basis is that the set D contain two elements at least; if D has only one element d it is then sufficient (but not quite necessary) that $C \cup \{\neg d\}$ be not interderivable with the set T of all true sentences. These results will be derived in section 2.

In section 3 I shall return to the problem of verisimilitude, and illustrate the usefulness of section 2's results in the treatment of it. In section 4 I shall discuss briefly some variants.

1 False theories are undoubtedly objectively distinct (but not necessarily distinguishable) from true theories. But is there any comparable objective difference amongst false theories? Are some false theories *closer to the truth* than others are; and if so, what is the objective determinant of their greater truthlikeness? This is the problem of *verisimilitude*.

The problem was first seriously discussed by Popper.[2] His resolution of the problem depended on bisecting the consequence class, or content, of any deductive system A into its true consequences, or *truth content*, $Ct_T(A)$, and its false consequences, or *falsity content*, $Ct_F(A)$. He proposed that—and I shall call this condition ($\Diamond$)—a theory B is closer to the truth than is a theory A if and only if (a) $Ct_T(A) \subset Ct_T(B)$ and $Ct_F(B) \subseteq Ct_F(A)$ or (b) $Ct_T(A) \subseteq Ct_T(B)$ and $Ct_F(B) \subset Ct_F(A)$.[3] Recently this proposal has been shown to be inadequate in Tichý's [1974] and in my [1974]. What Tichý and I showed independently was that if by these canons ($\Diamond$) the theory B is closer to the truth than the theory A is, then B must be true. Thus no two false theories are comparable by verisimilitude. A somewhat stronger result, that if the system T of all true sentences is unaxiomatisable then amongst axiomatisable systems only true ones are comparable by verisimilitude, was also derived in my [1974]. I shall later make use of a result close to this.

---

[1] Tarski [1930*a*], p. 35, and [1930*b*], p. 88.

[2] Popper [1963], Chapter 10 and addenda; [1966]; and [1972], Chapters 2 and 9.

[3] Popper [1963], p. 233; [1972], p. 52.

Here I shall say no more about the background to the problem of verisimilitude. The reader is referred to Popper's cited works, and to my [1974], the latter particularly for matters of terminology and notation. Instead I turn immediately to a modification of Popper's condition ($\Diamond$) which was in essentials suggested by Martin Hyland. I should like to make quite clear my indebtedness to Hyland for this suggestion, which is not only highly ingenious but, in view of the results of Tichý and myself, a suggestion fully in the spirit of Popper's original approach. As with Popper's own idea, the inadequacy of Hyland's proposal does not at all destroy its value.

Hyland's original modification was that, instead of comparing directly $Ct_T(A)$ and $Ct_T(B)$, and $Ct_F(A)$ and $Ct_F(B)$, we should take notice only of certain subsets of these sets; in particular, subsets $T_A \subseteq Ct_T(A)$, $T_B \subseteq Ct_T(B)$, $F_A \subseteq Ct_F(A)$, $F_B \subseteq Ct_F(B)$ such that $A = Cn(T_A \cup F_A)$ and $Cn(T_B \cup F_B) = B$. (If C is a set of sentences, $Cn(C)$ stands for the set of all the logical consequences of C.) We could then say that B was closer to the truth than A was if and only if either (a) $T_A \subset T_B$ and $F_B \subseteq F_A$ or (b) $T_A \subseteq T_B$ and $F_B \subset F_A$. But one thing stops this from being, as it stands, an attractive suggestion; that is, that trivially every system will be, according to what has been said, closer to the truth than itself.

This is the reason that I introduce the stricter requirement that $T_A \cup F_A$ and $T_B \cup F_B$ also be *bases* for A and B respectively. (This further requirement makes it quite obvious how the original idea leads to one theory's being closer to the truth than itself.) The proposal in full then becomes:

> ($\Diamond'$) a theory B is closer to the truth than is a theory A if and only if there are sets $T_A \subseteq Ct_T(A)$ and $F_A \subseteq Ct_F(A)$
> such that $T_A \cup F_A$ is a basis for A,
> and sets $T_B \subseteq Ct_T(B)$ and $F_B \subseteq Ct_F(B)$
> such that $T_B \cup F_B$ is a basis for B,
> and either (a) $T_A \subset T_B$ and $F_B \subseteq F_A$ or (b) $T_A \subseteq T_B$ and $F_B \subset F_A$.

So stated, the proposal is full of appeal. It can be said to recognise one crucial defect of Popper's ($\Diamond$), that it treats every element of the truth content or falsity content of a system on a par. This means that if we invite a new true sentence to join a false system we must simultaneously absorb into its content a large horde of 'gatecrashers', true and false, each of which is as fully a member of the truth content or falsity content as is the original guest. In consequence, as the truth content grows, so does the falsity content—and this is even true quantitatively, as shown in section 4 of my [1974]. What ($\Diamond'$) does, very crudely, is to allow only the man invited to count as a real guest, his attendant hangers-on being disregarded in the counting of true and false 'heads'.

Let us consider a simple example, given by Hyland, to see more formally what is intended. Let C be a set of true sentences, D a non-empty set of false sentences, and t some other true sentence, such that $C \cup \{t\} \cup D$ is an independent set. Let A be $Cn(C \cup D)$ and B be $Cn(C \cup \{t\} \cup D)$. One would like B to be closer to the truth than A is, since it differs from A only by containing an extra true 'axiom'. And indeed B *is* closer to the truth than A is, by the lights of ($\Diamond'$). We need only set $T_A = C$, $T_B = C \cup \{t\}$, $F_A = F_B = D$. Since D

is not empty[1] this means that the false system B is reckoned by $(\Diamond')$ to be closer to the truth than is the false system A. Thus, unlike condition $(\Diamond)$, condition $(\Diamond')$ is not obviously too narrow: it allows false theories to be compared for truthlikeness. Unfortunately, as I intend to show, it is too wide. For suitable juggling allows $(\Diamond')$ also to judge A closer to the truth than B is. *Every interesting ordering licensed by* $(\Diamond')$ *can be reversed.* Or, better, $(\Diamond')$ *does not define an ordering.*

To show the relevance to this conclusion of the results announced at the beginning of the paper I shall first discuss an example very similar to, but subtly different from, that given by Hyland. For notational convenience I shall introduce the symbolism $A \approx C :\cup: D$ to mean that $C \cup D$ is a basis for the *false* deductive system A, and that C comprises the true part of the basis, D the *non-empty* false part. In line with this I shall assume the sign $\cup$ to represent *disjoint* union. Now suppose that $G \approx J :\cup: \{f\} \cup K$ and $H \approx J :\cup: K$, so that G differs from H only by containing an extra false 'axiom' f. Clearly $(\Diamond')$ places H closer to the truth than it does G. We can put $T_G = T_H = J$; $F_G = \{f\} \cup K$, and $F_H = K$. On the assumption that K is not empty[2] the false system H is reckoned closer to the truth than is the false system G. But to reverse this decision we need only choose some $k \in K$. Then, as we shall check, $G \approx J \cup \{k \rightarrow f\} :\cup: K$ and $H \approx J :\cup: K$. It follows from this and $(\Diamond')$ that G is closer to the truth than H is.

The only aspect really needing proof is that $J \cup \{k \rightarrow f\} \cup K$ is an independent set. Note first that if we had $J \cup K \Vdash k \rightarrow f$ we would also have $J \cup K \Vdash f$, which is not possible, since $G \approx J :\cup: \{f\} \cup K$.[3] And secondly, if for some x in $J \cup K$ we had $J \cup \{k \rightarrow f\} \cup K \setminus \{x\} \Vdash x$ then, since $f \Vdash k \rightarrow f$, we would also have $J \cup \{f\} \cup K \setminus \{x\} \Vdash x$, which is again impossible. Thus $G \approx J \cup \{k \rightarrow f\} :\cup: K$.

No such simple manoeuvre will be uniformly successful in Hyland's example, where $A \approx C :\cup: D$ and $B \approx C \cup \{t\} :\cup: D$. The natural thing to try is conjoining the sentence t with some false element d of D. We could then show that B is the consequence class of the set $C \cup \{t \,\&\, d\} \cup D$. But this set is not a basis for

---

[1] Outside this footnote we shall always assume B false. For imagine it true. According to $(\Diamond)$—as I show in my [1974]—B is closer to the truth than A is if and only if A's truth content is a subsystem of B (a very trivial result). This provides little scope for serious comparisons. The condition $(\Diamond')$, on the other hand, is far too generous: *a true B is closer to the truth than is any false A whatever.* (Of course, by $(\Diamond)$ or $(\Diamond')$ no false system is closer to the truth than is any true system. But this does not mean that a true system, like a tautology, should be closer to the truth than any false one, such as Newton's theory.) The proof is very easy. For clearly we will have $F_B \subset F_A$, so that the condition reduces to $T_A \subseteq T_B$. Set, therefore, $T_A = \varnothing$, and take for F_A some basis for A, all of whose elements are false. This can always be done, as is obvious from the last paragraph of the proof of Theorem 4 below. If T_B is a basis for B, the proof is complete. (I rely here on Theorem 17 of Tarski [1930*a*], which asserts that every set of sentences has a basis.)

This result on its own indicates weaknesses in $(\Diamond')$. But, more importantly, it allows us in the text to restrict ourselves to the more interesting case of assuming the system in the 'more truthlike' position—and therefore the other one—to be false.

[2] We have made this assumption in our notation $H \approx J :\cup: K$. See also the previous note.

[3] The expression $B \Vdash A$ means what is normally meant by $B \vdash A$, an expression I reserve for unilateral entailment. See section 2 of my [1974].

B—both t & d and d belong to it. Moreover, we cannot simply discard d, even though $B \approx C :\cup: \{t \,\&\, d\} \cup D \setminus \{d\}$. For in this case the false part of A's basis is not contained in B's. Nor can we make use of the sentence $d \rightarrow t$; for this is true. But the problem can be solved in almost all cases, as will appear from THEOREM 4 below.

It can be seen that there are two distinguishable problems here. One, which is largely of logical—or perhaps metalogical—interest, was stated in two halves at the outset of the paper, and concerns the possibilities that exist for transferring the true part of a theory's basis to its false part, and vice versa. Symbolically: if $A \approx C :\cup: K \cup D$ and $B \approx C' \cup J :\cup: D'$, where J and K are both not empty, under what conditions are there sets J' and K' such that $A \approx C \cup K' :\cup: D$ and $B \approx C' :\cup: J' \cup D'$? We will attack this problem immediately in 2. The other problem, more directly relevant to Hyland's proposal ($\Diamond'$), will be solved in 3. It can be stated thus: if $A \approx C :\cup: K \cup D$ and $B \approx C \cup J :\cup: D$, where J and K are not both empty, under what conditions are there sets C', D', J', K' such that $A \approx C' \cup K' :\cup: D'$ and $B \approx C' :\cup: J' \cup D'$? Clearly, if the answer to the first were 'Always' so would be the answer to the second. But even though this is not the case, the solution to the one problem renders considerable assistance in the solution of the other.

2 We take the easy half of the logical problem first.

THEOREM 1: Suppose that $A \approx C :\cup: K \cup D$. Then for there to be a set K' such that $A \approx C \cup K' :\cup: D$ it is necessary and sufficient that $D \neq \varnothing$.

Proof. The necessity of the condition is immediate. A false theory cannot have a true basis. Assume then that D is not empty, and choose some $d \in D$. If K is empty, put $K' = \varnothing$. Otherwise, suppose the elements of K to be somehow well-ordered, lexicographically perhaps, as $k_0, k_1, \ldots$. As shown above, $A \approx C \cup \{d \rightarrow k_0\} :\cup: K \setminus \{k_0\} \cup D$. Likewise, $A \approx C \cup \{d \rightarrow k_0,\ d \rightarrow k_1\} :\cup: K \setminus \{k_0, k_1\} \cup D$. Continue transferring the elements $k_0, k_1, \ldots$ of K until K is exhausted. At every stage we will have a basis for A. We now show that $A \approx C \cup \{d \rightarrow k \mid k \in K\} :\cup: D$. Were there any x belonging to this purported basis that was derivable from the remainder, so would it be from some finite subset of the remainder. Thus there would be a natural number n (possibly depending on the sentence x) such that $C \cup \{d \rightarrow k_m \mid m < n\} \cup D$ would not be independent. Consequently, $C \cup \{d \rightarrow k_m \mid m < n\} \cup K \setminus \{k_m \mid m < n\} \cup D$ would not be independent either, contrary to the construction. The equivalence of A with $C \cup \{d \rightarrow k \mid k \in K\} \cup D$ is immediate.

Thus all but one of the false sentences in a false system's basis may be exchanged for true sentences without otherwise dislodging the basis. There is no such simple way, however, of replacing true sentences by false ones. We have already had a suggestion of this, but it is made quite plain by the following consideration. Suppose that the set T of all true sentences in the language is finitely axiomatisable, and therefore axiomatisable by a single sentence t_0. (This is not an assumption that I should care to make normally. But we are here only concerned with an illustration.) Let t_1 be any true sentence, and B be the false theory whose basis is $\{t_1, \neg\, t_0\}$. (It is easily seen that if f is false then the pair $\{f \vee t_0, \neg\, t_0\}$ is a basis for f. Thus B could be the consequence class of any false sentence.) We have $B \approx \{t_1\} :\cup: \{\neg\, t_0\}$. But there is no basis for B that contains, along with $\neg\, t_0$, any other false sentence. For $\neg\, t_0$ is derivable from every false sentence. More generally, if $B \approx J :\cup: \{\neg\, t_0\}$ for some independent set of true sentences J there will be no basis for B containing $\neg\, t_0$ and other false sentences. Thus no true sentences can be turned into false ones in any such case.[1]

From a particular instance we return to the general problem, almost stated above as: if $B \approx C \cup J :\cup: D$, under what conditions is there a set J' such that $B \approx C :\cup: J' \cup D$? We know that D must not consist of a single element $\neg\, t_0$ where t_0 is an axiomatisation of T. Are there any other conditions under which no such set J' can be found? Indeed there are, conditions which are a simple generalisation of those already encountered (which can be regarded as the special case of $C = \varnothing$). These conditions are spelt out in the next two theorems, the first of which merely clears the way for the second.

THEOREM 2: Suppose that $B \approx C \cup J :\cup: D$. Then if D has more than one element there exists a set J' such that $B \approx C :\cup: J' \cup D$.

Proof. Let D have two elements d_0 and d_1. If $J = \varnothing$ put $J' = \varnothing$. Otherwise let j_0 be some element of J. Then $B \approx C \cup J \setminus \{j_0\} :\cup: \{j_0 \,\&\, (d_0 \vee d_1)\} \cup D$.

The equivalence with B of this purported basis is trivial. So we need only check independence. First, suppose that $C \cup J \setminus \{j_0\} \cup D$ entails $j_0 \,\&\, (d_0 \vee d_1)$; it must then entail j_0, which conflicts with the independence of $C \cup J \cup D$. Second, suppose that for some x, not j_0, in $C \cup J \cup D$ we had $C \cup J \setminus \{j_0\} \cup \{j_0 \,\&\, (d_0 \vee d_1)\} \cup D \setminus \{x\} \Vdash x$. Then whatever x may be, d_0 or d_1 or some other sentence, we will have $C \cup J \cup D \setminus \{x\} \Vdash x$, which again conflicts with $C \cup J \cup D$'s independence. This completes the proof.

Thus one true element j_0 of J can be eliminated in favour of a false sentence $j_0 \,\&\, (d_0 \vee d_1)$. By lexicographically ordering the elements of J we can, as in THEOREM 1, transfer them all. The proof that the resulting set $C \cup \{j \,\&\, (d_0 \vee d_1) \,/\, j \in J\} \cup D$ is independent relies, as before, on the fact that all proofs are finitely long.

We now consider only the case where D has a single element d. Indeed, it is enough to restrict ourselves to a set J with only one element. For if we can transfer just one true sentence in $C \cup J$ to join d, the basis so obtained will have two false elements; and so THEOREM 2 will become applicable.

THEOREM 3: Suppose that $B \approx C \cup \{j\} :\cup: \{d\}$. Then for there to be a sentence j' such that $B \approx C :\cup: \{j'\} \cup \{d\}$ it is necessary and sufficient that $C \cup \{\neg\, d\}$ be not interderivable with the set T of all true sentences.

Proof. Since $C \cup \{j'\} \cup \{d\}$ is to be a basis for B, and $B \approx C \cup \{j\} :\cup: \{d\}$, the sentence j' will have to be weaker than $j \,\&\, d$, given C, yet strong enough for $j \,\&\, d$ and $j' \,\&\, d$ to be equivalent, given C. In fact, as we shall show, it is enough if j' is a simple consequence of $j \,\&\, d$.

---

[1] For an application of this example see THEOREM 4 *below*; in a way, the example sums up all the difficulties of our problem.

If it is, then it will be of the form $(j \,\&\, d) \vee f$, where f is a false sentence yet to be determined. Our requirements on f are then (1) that it is false; (2) that $C \cup \{((j \,\&\, d) \vee f) \,\&\, d\}$, which is equivalent to $C \cup \{(j \,\&\, d) \vee (f \,\&\, d)\}$, be also equivalent to $C \cup \{j \,\&\, d\}$; therefore that $C \cup \{f \,\&\, d\} \Vdash j$; (3) that not $C \cup \{(j \,\&\, d) \vee f\} \Vdash d$; (4) that not $C \cup \{d\} \Vdash (j \,\&\, d) \vee f$; and (5) that no element x of C is derivable from $C \setminus \{x\} \cup \{(j \,\&\, d) \vee f\} \cup \{d\}$ (but this will hold whatever f is, and is easy to prove; so I shall not prove it).

Requirement (3) is the same as saying that not $C \cup \{f\} \Vdash d$; and (4) is the same as saying that not $C \cup \{\neg f \,\&\, d\} \Vdash j$. But since, by (2), $C \cup \{f \,\&\, d\} \Vdash j$, we can dispense with requirement (4); it is superfluous. For if as well as (2) we had $C \cup \{\neg f \,\&\, d\} \Vdash j$, we would have $C \cup \{d\} \Vdash j$, which is impossible, since $C \cup \{j\} \cup \{d\}$ is a basis for B. Thus the conditions boil down to three: that f be false, that $C \cup \{f \,\&\, d\} \Vdash j$, and that not $C \cup \{f\} \Vdash d$.

Suppose that these conditions are never met. Then for any false f, if $C \cup \{\neg j\} \cup \{d\} \Vdash \neg f$ then $C \cup \{\neg d\} \Vdash \neg f$. Or, writing e for $\neg f$: if $T \Vdash e$ and $C \cup \{\neg j\} \cup \{d\} \Vdash e$ then $C \cup \{\neg d\} \Vdash e$. In the terminology of my [1974], this amounts to saying that the truth content of the deductive system $Cn(C \cup \{\neg j \,\&\, d\})$ is included within the content of the system $Cn(C \cup \{\neg d\})$. Symbolically (see [1974], section 2),

$$Cn(C) \,\&\, \neg d \Vdash (Cn(C) \,\&\, (\neg j \,\&\, d)) \vee T.$$

By THEOREM 5 of that paper, since the system $Cn(C) \,\&\, \neg d$ is true, and, trivially, $Cn(C) \,\&\, (\neg j \,\&\, d) \Vdash d$, which is false, we can conclude that $Cn(C) \,\&\, \neg d \,/\!/\, T$. That is to say, the set $C \cup \{\neg d\}$ is interderivable with the set T of all true sentences.

In consequence, if $C \cup \{\neg d\}$ is not interderivable with T, there is a sentence f with the stated properties, so there is a sentence j', namely $(j \,\&\, d) \vee f$, that satisfies $B \approx C :\!\cup\!: \{j'\} \cup \{d\}$.

Conversely, suppose that $C \cup \{\neg d\}$ is interderivable with T. If j' is any false sentence $C \cup \{\neg d\} \Vdash \neg j'$, so $C \cup \{j'\} \Vdash d$. Thus no set $C \cup \{i'\} \cup \{d\}$ can be independent if j' is false.

COROLLARY 1: Suppose that $B \approx C \cup J :\!\cup\!: \{d\}$, and that $J \neq \varnothing$. For there to be a set of sentences J' such that $B \approx C :\!\cup\!: J' \cup \{d\}$ it is necessary and sufficient that $C \cup \{\neg d\}$ be not interderivable with the set T of all true sentences.

Proof. The condition is certainly sufficient, in view of THEOREM 3 and THEOREM 2. That it is necessary is proved just as the necessity half of THEOREM 3 was proved.

COROLLARY 2: Suppose that $B \approx C \cup J :\!\cup\!: \{d\}$, and that $J \neq \varnothing$. Let T not be finitely axiomatisable. Then if B is finitely axiomatisable there exists a set of sentences J' such that $B \approx C :\!\cup\!: J' \cup \{d\}$.

Proof. Since all proofs are finitely long, the proof of the sentence B from $C \cup J \cup \{d\}$ can involve only finitely many of these sentences. Thus there is a finite subset of $C \cup J \cup \{d\}$ that is interderivable with B. But no proper subset of an independent set can be interderivable with the whole set. Thus C is finite, in which case if $C \cup \{\neg\, d\}$ were interderivable with T it would be a finite axiomatisation of it. It follows that a set J' exists.

This concludes my discussion of the problem stated at the start of the paper. I turn now to its application to the proposal $(\Diamond')$.

3 The following THEOREM provides the solution to our second problem.

THEOREM 4: Suppose that $A \approx C :\cup: K \cup D$ and $B \approx C \cup J :\cup: D$, where J and K are not both empty. Then for there to be sets C', D', J', K' such that $A \approx C' \cup K' :\cup: D'$ and $B \approx C' :\cup: J' \cup D'$ it is necessary and sufficient that A not be an irreducible theory.

Proof. To show that the condition is necessary, suppose that A is irreducible. Since it is false, $C \cup K \cup D$ contains only one element, which must be $\neg\, t_0$, where t_0 is interderivable with T (see the footnote on p. 182 above). But $B \approx C \cup J :\cup: D$, so D is not empty. Thus $K = \varnothing$ and $C = \varnothing$. In other words, $A \approx \varnothing :\cup: \{\neg\, t_0\}$ and $B \approx J :\cup: \{\neg\, t_0\}$, where J is not empty. Now it is easily seen that A has no other basis, in that $\neg\, t_0$ cannot be weakened and must not be strengthened. And there is no set J' such hat $B \approx \varnothing :\cup: J' \cup \{\neg\, t_0\}$. Thus there are no new bases of the right sort.

We now turn to the sufficiency of the condition. From THEOREMS 1, 2, and 3 we know that sets C', D', J', K' exist as required if D has more than one element; and also if D has a solitary element d and $C \cup \{\neg\, d\}$ is not interderivable with T. For we can simply take $C' = C$ and $D' = D$ and transfer the bases' elements as before.

Suppose that D has one element d only. I shall show that if $C \neq \varnothing$ there are sets C', D', J', K' as required. If $C = \varnothing$ and $C \cup \{\neg\, d\}$ is interderivable with T, then A is irreducible. This case therefore need not concern us.

We have $A \approx C :\cup: K \cup \{d\}$ and $B \approx C \cup J :\cup: \{d\}$. By THEOREM 1 we can write $A \approx C \cup \{d \rightarrow k \mid k \in K\} :\cup: \{d\}$. The simplest way of proceeding now is to put $C' = \varnothing$, $D' = \{c \,\&\, d \mid c \in C\}$, $K' = \{d \rightarrow k \mid k \in K\}$, and $J' = \{j \,\&\, d \mid j \in J\}$. Since C is not empty, neither is D'. Moreover, $A \approx C' \cup K' :\cup: D'$ and $B \approx C' :\cup: J' \cup D'$. The proofs of these facts are very straightforward and, indeed, similar to previous proofs in the paper. They are therefore omitted.

What moral can we draw from THEOREM 4? Clearly a devastating one as far as $(\Diamond')$ is concerned. For in our new notation this condition proposes that B is closer to the truth than A is if and only if there are sets C, D, J, K, of which J and K are not both empty, such that $A \approx C :\cup: K \cup D$ and $B \approx C \cup J :\cup: D$. Thus it follows at once from THEOREM 4 that unless A is irreducible, if B is closer to the truth than A is, then A is closer to the truth than B is. (If A is irreducible and false—only a possibility if T is finitely axiomatisable—then any other false theory is closer to the truth than A is.) As I asserted above, $(\Diamond')$ *does not define an ordering.*

4 (*Added December 1973*): The preceding three sections were written in August 1973. Since then, conversations with Professor Popper and correspondence with Dr Tichý and with Professor Zoltan Domotor have encouraged me to develop the investigation a little further. For, despite the appearance of utter collapse, there is perhaps still some room for manoeuvring around ($\Diamond'$). In this section I shall briefly review some of the possibilities.

The most obvious one stems from the asymmetry between THEOREM 1 and the combination of THEOREMS 2 and 3: the fact that false sentences can be exchanged for true ones without difficulty, whereas the reverse is sometimes not the case. In a few moments I shall show how this balance can be righted. Till then, can we exploit it?

We might call B closer to the truth than is A if and only if there are sets C, D, J, K such that $A \approx C :\cup: K \cup D$ and $B \approx C \cup J :\cup: D$, but no sets J', K' such that $A \approx C \cup K' :\cup: D$ and $B \approx C :\cup: J' \cup D$. It is not obvious that we have here defined an ordering, but we shall show at once that this is in fact so.

For THEOREMS 1, 2, 3 allow us to reduce the conditions to the existence of C, K, d, and non-empty J such that $A \approx C :\cup: K \cup \{d\}$, $B \approx C \cup J :\cup: \{d\}$, and $C \cup \{\neg d\} // T$. But as shown in the last paragraph of the proof of THEOREM 3, this last condition entails the emptiness of K. Thus for B to be more truthlike than A is we must have $A \approx C :\cup: \{d\}$ and $B \approx C \cup J :\cup: \{d\}$. In other words, a necessary condition for the false theory B to be nearer the truth than is the false theory A is that $B \vdash A$. This is an ordering anyway, and cannot be reversed.

However, it has some odd features. In particular, neither A nor B is permitted to be finitely axiomatisable, as was proved in COROLLARY 2 to THEOREM 3. Thus *no finitely axiomatisable systems fall within the scope of this theory of truthlikeness.* Just as serious, however, is the following. Although we may be able to tolerate a verisimilitude ordering which, amongst false theories, is even more restrictive than their ordering by $\vdash$ (for it may provide us with some genuine distinctions —see my [1974], p. 168), we must certainly disallow an ordering that *coincides* with $\vdash$ (*op. cit.*, p. 167). This is of course false of this ordering, for not every C, d satisfy $C \cup \{\neg d\} // T$. But even so, this ordering may be too like $\vdash$ for comfort. It would be most unsatisfactory, in particular, if $B \vdash A$, together with $A \approx C :\cup: \{d\}$ and $C \cup \{\neg d\} // T$ entailed that B was closer to the truth than A was. But I do not know if this is so. In other words, the following problem is, as far as I know, open: if $B \vdash A$ and $A \approx C :\cup: \{d\}$, where $C \cup \{\neg d\} // T$, is there a set J such that $B \approx C \cup J :\cup: \{d\}$? More generally, suppose that neither A nor B is finitely axiomatisable, and that $B \vdash A$. Can every basis for A be extended to a basis for B? Alternatively, is there a basis for A that can be extended to a basis for B?[1]

I turn now to ways of tampering with ($\Diamond'$) by imposing conditions stronger than mere independence. The first of these requires that the bases be also *maximally independent*.

---

[1] As pointed out in footnote 1 on p. 180 *above*, every system has a basis. THEOREM 19 of Tarski [1930*a*] shows that a finitely axiomatisable system only has finite bases. (Only finitely many sentences can be needed to derive finitely many sentences.) From this it is evident that if A is finitely axiomatisable and $B \vdash A$ then we can extend every basis for A to a basis for B; and, at the other extreme, if A is not finitely axiomatisable, but B is, there will be no hope of extending a basis for A to a basis for B.

Thus only the case of neither A nor B finitely axiomatisable is likely to cause any difficulty. And, indeed, in the immediate context it is the only case of any real interest.

A set C of sentences is maximally independent (Sheffer [1921], pp. 32f.), or a *maximal basis*, if for any two distinct elements c_0, c_1 of C we have $\Vdash c_0 \vee c_1$. (This is of course in addition to the demand that C be a basis.) Every system has a maximal basis; see the previous footnote and Tarski [1935], p. 362. It is immediately evident that, if C is maximally independent, it contains at most one false element—the sentence $c_0 \vee c_1$ will not be demonstrable if each disjunct is false. Thus if we modify (◇′) by requiring that the bases be maximal bases, we shall have $A \approx C$:U: $\{d\}$ and $B \approx C \cup \mathcal{J}$:U: $\{d\}$, where $\mathcal{J} \neq \varnothing$ and the notation is taken to imply maximality. Thus, as above, $B \vdash A$.

Here, however, it is not said that $C \cup \{\neg\, d\}$ must be interderivable with T. So finitely axiomatisable theories can stand in the relation offered by (◇′); and, in fact, the results reported in the previous footnote still go through. In particular, *if A is finitely axiomatisable, every stronger theory*—even the contradictory one—*is nearer to the truth*. And *if A is not finitely axiomatisable, no finitely axiomatisable theory is nearer to the truth*. These are almost as counterintuitive as the restrictions encountered earlier.

So much for maximal independence. What about complete independence? This is an idea originally introduced by E. H. Moore ([1910], p. 82). Moore called a set of m properties *completely independent* if every one of the 2^m Q-properties (as Carnap [1950], pp. 122–6, called them) defined from them is instantiated. Thus the holding of some of the properties neither requires nor precludes the holding of others. Analogously, a set of sentences may be called completely independent, or a *complete basis* if and only if it is consistent to suppose any subset of the set true, the remainder false. Thus C is a complete basis if for every $C_0 \subseteq C$ there is an interpretation under which all the elements of C_0 hold, while all those of $C \setminus C_0$ fail. (Thus complete independence involves consistency.) Simple independence is the result of limiting $C \setminus C_0$ to one element.

The analogues for Theorems 1–3 are sufficiently significant for us to state them in a further Theorem. For this purpose we shall employ the notation $A \sim C$:U: D to signify that $C \cup D$ is a complete basis for the false theory A, C all being true sentences, D false ones. It is not obvious to me that every theory has a complete basis. But if a false theory does have one with more than a single element, then it has others, as Theorem 5 will show. The proof is a straightforward extension of my criticism of Tichý in section 6 of my [1974].

Theorem 5: Suppose that $A \sim C$:U: $K \cup D$ and $B \sim C \cup \mathcal{J}$:U: D, where $\mathcal{J}$, K are not both empty. Then there exist sets $\mathcal{J}'$, K' such that $A \sim C \cup K'$:U: D and $B \sim C$:U: $\mathcal{J}' \cup D$.

Proof. By assumption D is not empty, since B is false. Choose $d \in D$. Then we write $\mathcal{J}' = \{d \leftrightarrow j \mid j \in \mathcal{J}\}$, $K' = \{d \leftrightarrow k \mid k \in K\}$. The falsity of d ensures that the sentences in $\mathcal{J}'$ are all false, whilst those in K' are all true.

The equivalences of $C \cup K' \cup D$ with A and $C \cup \mathcal{J}' \cup D$ with B are immediate.

It remains only to prove that the sets $C \cup K' \cup D$ and $C \cup \mathcal{J}' \cup D$ are each completely independent. One proof suffices; so we will establish complete independence for $C \cup K' \cup D$.

We select an arbitrary subset of $C \cup \{d \leftrightarrow k \mid k \in K\} \cup D$ and show that there is an interpretation $\mathfrak{A}$ in which all its elements hold, whilst the remainder fail. Call the subset $E \cup \{d \leftrightarrow k \mid k \in K_0\}$, where $E \subseteq C \cup D$ and $K_0 \subseteq K$. Everything now depends on whether $d \in E$.

If $d \in E$, choose $\mathfrak{A}$ such that $E \cup \{k \mid k \in K_0\}$ hold and $((C \cup D) \setminus E) \cup \{k \mid k \in K \setminus K_0\}$ fail. This is always possible, because $C \cup K \cup D$ is a completely independent set, and the two sets mentioned are complementary subsets of it. Since d holds in $\mathfrak{A}$, $d \leftrightarrow k$ will hold in $\mathfrak{A}$ if and only if k holds; thus $E \cup \{d \leftrightarrow k \mid k \in K_0\}$ hold in $\mathfrak{A}$, and the remaining members of $C \cup K' \cup D$ do not.

On the other hand, if $d \notin E$ we will want d to fail in the structure $\mathfrak{A}$. If it does, $d \leftrightarrow k$ will hold in $\mathfrak{A}$ if and only if k fails. We therefore choose $\mathfrak{A}$ so that $E \cup \{k \mid k \in K \setminus K_0\}$ all hold, while the rest of $C \cup K \cup D$ do not. Again this can always be done; and again it is easy to see that the structure chosen does the trick.

Thus $C \cup K' \cup D$ is a completely independent set of sentences.

It follows from this Theorem that there is no future in modifying $(\Diamond')$ by writing 'complete basis' for 'basis'. Indeed, to the extent that it is possible to regress from a position as gravely unsatisfactory as $(\Diamond')$, the modification would be retrograde.

What the Theorem does show positively is how the asymmetry apparent in Theorems 1, 2, and 3 can be righted. It is clear, moreover, that if we had, mimicking the hypothesis of Theorem 3, $B \sim C \cup \{j\} :\cup: \{d\}$, we would not have $C \cup \{\neg d\} \Vdash j$, and therefore not $C \cup \{\neg d\} // T$.

In conclusion I should like to mention *organic* independence. An axiom is *organic with respect to a system* A if it contains no segment which is a theorem of the system A 'or becomes one as soon as its variables have been bound with an appropriate quantifier'.[1] A basis is organic if all its elements are organic. Would there be any point in restricting $(\Diamond')$ to organic bases?

So little seems to be known about organic bases (as a measure of how little, see the Index of Reviews by Subjects in *The Journal of Symbolic Logic*, **26**, 1961, and later volumes; the heading is 'organic axioms'), that it is almost impossible to answer this question. To be sure, the constructions in Theorems 1–5 above generally yield inorganic bases, so that little of our criticism would be effective in the event of such a modification. But, on the other hand, unless we know what variety of theories have organic bases, we run the risk of restricting the scope of this theory of truthlikeness to the point of pointlessness. Yet even if every theory possesses an organic basis of some description, what raises the confinement of $(\Diamond')$ to organic bases above the level of a purely *ad hoc* manoeuvre? It may be that organic bases have a privileged status, perhaps as exhibiting best the 'unity' of the systems that they underlie. But this is not obvious, and surely deserves a morsel of critical discussion.

---

[1] This is a combination of the definition, due to Wajsberg, given in Łukasiewicz and Tarski [1930] and the definition, due to Wajsberg, given in Sobociński [1955–1956]. It is hard to see what makes a quantifier appropriate; the existential quantifier certainly does not seem appropriate if our language contains the sign of identity.

In short, I do not see any of the possibilities inspected in this section as live possibilities. But this is not to deny that they *are* possibilities.

DAVID MILLER
University of Warwick

REFERENCES

CARNAP, R. [1950]: *Logical Foundations of Probability.*

ŁUKASIEWICZ, J. and TARSKI, A. [1930]: 'Untersuchungen über den Aussagenkalkül', *Comptes Rendus des séances de la Société des Sciences et des Lettres de Varsovie*, **23**, cl. iii, pp. 30–50. References are to the translation by J. H. Woodger in Tarski [1956].

MILLER, D. [1974]: 'Popper's Qualitative Theory of Verisimilitude', *British Journal for the Philosophy of Science*, **25**, pp. 166–77.

MOORE, E. H. [1910]: 'Introduction to a Form of General Analysis', *The New Haven Mathematical Colloquium*, pp. 1–150.

POPPER, K. R. [1962]: 'Some Comments on Truth and the Growth of Knowledge', in E. Nagel, P. Suppes, and A. Tarski (*eds.*): *Logic, Methodology and Philosophy of Science*, pp. 285–92.

POPPER, K. R. [1963]: *Conjectures and Refutations.*

POPPER, K. R. [1966]: 'A Theorem on Truth-Content', in P. K. Feyerabend and G. Maxwell (*eds.*): *Mind, Matter and Method*, pp. 343–53.

POPPER, K. R. [1972]: *Objective Knowledge.*

SHEFFER, H. M. [1921]: 'The General Theory of Notational Relativity', mimeographed paper.

SOBOCIŃSKI, B. [1955–1956]: 'On Well Constructed Axiom Systems', *Rocznik Polskiego Towarzystwa Naukowego Na Obczyźnie*, **6**, pp. 54–63, 65f., 70.

TARSKI, A. [1930*a*]: 'Über einige fundamentale Begriffe der Metamathematik', *Comptes Rendus des séances de la Société des Sciences et des Lettres de Varsovie*, **23**, cl. iii, pp. 22–9. References are to the translation by J. H. Woodger in Tarski [1956].

TARSKI, A. [1930*b*]: 'Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften. I', *Monatshefte für Mathematik und Physik*, **37**, pp. 361–404. References are to the translation by J. H. Woodger in Tarski [1956].

TARSKI, A. [1935]: 'Grundzüge des Systemenkalkül, Erster Teil', *Fundamenta Mathematicae*, **25**, pp. 503–26. References are to the translation by J. H. Woodger in Tarski [1956].

TARSKI, A. [1956]: *Logic, Semantics, Metamathematics.*

TICHÝ, P. [1974]: 'On Popper's Definitions of Verisimilitude', *British Journal for the Philosophy of Science*, **25**, pp. 155–60.

Reviews

AJDUKIEWICZ, K. [1973]: *Problems and Theories of Philosophy.* (Translated by H. Skolimowski and A. Quinton.) Cambridge: Cambridge University Press. £2.90. Pp. 170.

This concise introduction to philosophy by Kazimierz Ajdukiewicz (1890–1963), the most outstanding of modern Polish philosophers, gives a lucid and most instructive survey of the classical problems and doctrines in the theory of knowledge and in metaphysics.

The first of its two parts, concerned with the theory of knowledge, discusses 'The problem of truth' (classical and non-classical definitions, scepticism and its refutation, the relation between non-classical definitions of truth and idealism), 'Problems of the sources of knowledge' (the psychological and epistemological versions of the problem, varieties of empiricism and of apriorism, the disputes between empiricism and apriorism concerning the nature of mathematical statements, conventionalism, rationalism versus irrationalism) and 'The problem of the limits of knowledge' (immanent and transcendental epistemological idealism, realism, positivism, neo-positivism). The second part of the book, concerned with metaphysics, contains chapters on 'Ontology', 'Metaphysical inferences from reflections on knowledge' (The problem of ideal objects, the problem of metaphysical idealism and realism), 'Metaphysical problems arising from the investigation of nature' (the substance and structure of the world, soul and body, determinism versus indeterminism, mechanism versus finalism), 'Metaphysical problems arising from religion' and 'Metaphysics as the attempt to achieve an ultimate world-outlook'.

The first Polish edition of the book appeared in 1949. The text had been originally intended as a new introduction to the planned re-issue of a selection of the classics of philosophy (epistemology, logic and metaphysics) edited by Ajdukiewicz and published twenty-five years previously. The re-issue of that anthology never occurred. Instead the introduction appeared as a separate book and for years afterwards served students in Polish universities as a basic textbook for an introduction to philosophy course taken by thousands of them as part of their B.A. or B.SC. degrees. Though it naturally antedated many post-war developments and revivals in philosophy, it has suffered little from the passage of time. It was a good idea to make it available to English-speaking readers. A cheaper paperback edition is most advisable.

The author's intention was to show how problems and their solutions have been developing in philosophy, how modern problems and doctrines grew out of ancient ones (hence the title of the original 'Problems and trends in philosophy'; the Polish phrase 'kierunki filozofii' may be translated either as 'schools', 'doctrines' or 'trends in philosophy'). He used the traditional terminology and formulations whenever these were sufficiently simple and clear, providing his own clarifications for some of the more obscure or ambiguous doctrines. Some of these clarifications, though admittedly controversial, are original, illuminating and will be appreciated more by mature students or by philosophy teachers than by beginners. This applies, I think, to Ajdukiewicz's formulations of the positions of radical and moderate empiricism, immanent and transcendental epistemological idealism, conventionalism, positivism, objective metaphysical idealism, *etc.* To that extent the book is more than a standard, popular introduction to philosophy.

Although the author's own position transpires unmistakably from the text—it is that of moderate empiricism and conventionalism in epistemology and critical realism in metaphysics—he was at pains to give a fair, impersonal account of all doctrines and a lot of his effort went into presenting in an understanding fashion those doctrines of which he obviously disapproved, *e.g.* Kant's and Husserl's apriorism, objective idealism, dialectic materialism, *etc.* The obscurities and other inadequacies of traditional disputes are pointed out but without branding them as meaningless. This reflects both Ajdukiewicz's policy to use 'sympathetic understanding' as a basis of the logical analysis of traditional philosophical problems and doctrines and the relatively liberal pre-war tradition of Polish philosophy in which several rival schools interacted; at a typical philosophical conference in those days one would meet—apart from the dominant logically-minded philosophers of the Lwow and Warsaw schools—a leading phenomenologist (R. Ingarden), Thomists, Hegelians (followers of Hoene-Wronski) and several 'unattached' philosophers, one of whom was a very unorthodox marxist (Chwistek). Of course, the appearance of marxism as the philosophy of the post-war establishment in Poland, had also to be taken into account in Ajdukiewicz's book. In effect, one finds in it a wider spectrum of philosophical views than in any comparable book published at the time in the West (existentialism and intuitionism were not yet popular).

The translators of the book, Messrs Skolimowski and Quinton, have provided in their Preface (of 10 pages) a useful short note on 'philosophy and logic in modern Poland', Ajdukiewicz's biography with a list of some of his writings and a brief account of his philosophical views. This will be undoubtedly appreciated by readers unfamiliar with Ajdukiewicz's philosophy and his background.

I should like, however, to dispute a few points in the translators' Preface:

(1) My main doubts concern the comparisons Skolimowski and Quinton make between certain views of Ajdukiewicz and of Quine in their otherwise quite justified efforts to show that Ajdukiewicz's philosophy is not only 'with-it' but perhaps even 'before-it'. So on p. xvi they say that Ajdukiewicz's claim about the existence of languages which are 'radically untranslatable into each other' and between which an epistemologist cannot decide as impartial arbiter, was an anticipation of 'Quine's conclusions about indeterminacy of translation', fundamental to Quine's philosophy of language.

I find this suggestion doubtful. On the face of it, at least, some of Quine's views on translatability of languages seem to contradict flatly Ajdukiewicz's radical conventionalism.

According to Quine's thesis of indeterminacy of translation there are always '. . . many ways of translating the component sentences [of a theory], essentially different individually . . .' which '. . . . would deliver the same empirical implications for the theory as a whole . . .'.[1] We are also assured by Quine that the indeterminacy of translation has little bearing on observation sentences.

---

[1] Quine [1969], p. 80.

'The equating of observation sentence of our language to an observation sentence of another language is mostly a matter of empirical generalisation . . . .'[1] It would appear from these quotations that, in Quine's view, sufficiently rich languages containing observation sentences are always intertranslatable albeit not in a unique manner, owing to the indeterminacy of translation of their abstract, theoretical parts and owing to the 'inscrutability of reference'.

Now, according to radical conventionalism there exist languages, definable in terms of their vocabularies and in terms of axiomatic, deductive and empirical meaning rules (governing the acceptance of sentences) which share no (logical or extralogical) expressions and which are, therefore, not intertranslatable. Some of those languages, moreover, are closed in the sense of having no intertranslatable conservative extensions (*i.e.* cannot be extended by enriching their vocabularies with new expressions without thereby altering the meaning of already existing expressions or without thereby splitting the language into several disconnected languages). In other words, according to Ajdukiewicz's radical conventionalism there are languages with observation sentences (governed by empirical meaning rules) which not only are not intertranslatable but cannot be enriched to become inter-translatable. The languages of Newtonian and Relativistic mechanics—so Ajdukiewicz suggested—are of this type.

Whether these views of Ajdukiewicz and of Quine are in fact mutually inconsistent depends, of course, on whether 'translation' and other crucial words occurring in them are used in the same way. Ajdukiewicz regards an empirical language as uniquely determined by its vocabulary and by its matrix (determined, in turn, by its meaning rules and by experiential data). Two languages L and L' are then said to be intertranslatable by the relation R if and only if R is a one-one relation which to each expression of L assigns an expression (of the same syntactical category) of L' in such a way that the matrix of L is transformed into the matrix of L' when every expression in the former matrix is replaced by its image. An expression E of L has the same meaning as another expression of E' or L' if there exists a relation R which makes L and L' intertranslatable and R holds between E and E'. The relation of sameness of meaning is an equivalence relation. 'Meaning' is then defined as the family of equivalence classes generated by the relation of sameness of meaning.[2] This approach to language and translation may seem essentially different from Quine's. On the other hand, Ajdukiewicz—just as Quine—takes dispositions to assent or dissent to uttered sentences under certain conditions as the basis for the reconstruction of a language, the rejection of certain 'obvious' sentences as the indication of 'changing the subject' and although—unlike Quine—he based his philosophy of language on a conception of meaning, meaning for him was an abstract property of expressions rather than a mental image associated with an expression. To conclude, either the position of the two authors on translatability and translation are very difficult to compare or else—if we disregard the complicating differences of approach—they seem to contradict one another. In either case it seems implausible to suggest without qualifications and a great deal of explanatory comments that one of them anticipated the views of the other in the given area.

It seems more plausible to me to suggest that Quine's doctrine of the indeterminacy of translation for theories coupled with the claim that observation sentences are not affected by this indeterminacy had been vaguely anticipated by Poincaré (*a*) in his view on the observational equivalence of alternative geometries used in physics, (*b*) in his view on the observational equivalence of alternative physical theories, couched in different metaphorical expressions, (*c*) in his polemic with LeRoy in which he maintained that all sufficiently rich human languages based on the same logic are intertranslatable up to their observational content.[1]

---

[1] *op. cit.* p. 89.

[2] Ajdukiewicz [1934], chapter 10.

On the other hand, Ajdukiewicz's radical conventionalism, partly inspired by LeRoy's extreme 'nominalism' (to use Poincaré's terminology),[2] anticipated some of the later philosophy of Wittgenstein (a-psychological conception of meaning, multiplicity of languages which develop and evolve in their own individual ways, inappropriateness of using one of those languages to pass judgments on the qualities of another, lack of external, epistemological or otherwise, criteria of evaluation, *etc.*) and—in an obvious way—the so-called 'incommensurability theses', 'proliferation of theories' and similar doctrines so much discussed in recent years.

(2) On p. xvii Skolimowski and Quinton say that after 1945 Ajdukiewicz 'was led to abandon the kind of conventionalism or linguistic relativism that has just been outlined, but he entertained the possibility of another doctrine familiar from the writings of Quine which takes the logical aspect of language, in Ajdukiewicz the sentences determined as acceptable by its axiomatic and deductive rules, to be as much accessible to the arbitrament of experience as are sentences of a straightforwardly empirical character. Like Quine he maintained that empirical testing bears not on single assertions, but on whole bodies of assertions among which the propositions of logic may be included . . . .'

Now, the quoted passage is not quite correct and is misleading. It is not correct in so far as it appears to say that after 1945 Ajdukiewicz entertained the possibility of the view according to which the sentences of our language governed by the axiomatic and deductive meaning rules ('the logical aspect of language', as the translators call it) are susceptible to empirical testing, in particular, may be revised as a consequence of empirical testing. This view Ajdukiewicz certainly never entertained, for it would have been absurd on his understanding of 'sentences governed by axiomatic meaning rules'. What he did in his article 'Logic and experience'[3] and afterwards also in 'The problem of empiricism and the notion of meaning',[4] was to argue that the view of radical empiricism, according to which all components of (scientific) knowledge, including logic, are empirical, and the view of moderate empiricism, according to which apart from the empirical components of knowledge there are *a priori* analytic ones (logical tautologies being among them), are not necessarily incompatible. They are not incompatible if either they are understood to refer to different types of languages or if they are understood as programmes (prescriptions about how to construct the language of science) rather than as descriptive claims about the language of existing science. In the former case, radical empiricism is understood as referring to the languages in which there are no sentences governed by the axiomatic meaning rules, the only rules in those languages being deductive and empirical ones which are not sufficient to yield any analytic sentences. In such languages,

---

[1] Poincaré [1902], [1905].

[2] Poincaré [1905], part 3.

[3] Ajdukiewicz [1947].

[4] Ajdukiewicz [1964].

'laws of logic' would be used as auxiliary assumptions in testing scientific hypotheses; the negative outcomes of experiments or observations might be then accommodated either by modifying some of the scientific hypotheses or by modifying some of the laws of logic. It should be clear from what I have just said that, unlike Quine, Ajdukiewicz did not reject (in 'Logic and experience') the analytic-synthetic distinction. On the contrary, he emphasised that both the everyday language which we use and the language of science and of mathematics seem to be based on axiomatic and deductive (the former two also on empirical) meaning rules and, therefore, contain analytic sentences. In conformity with his conventionalist view of languages, which he never completely rejected, he argued that it is nevertheless possible to construct languages without axiomatic meaning rules (and, therefore, without analytic sentences) and consequently that radical empiricism as a programme for constructing science is feasible.

The passage quoted from the translators' Preface is also historically misleading. For it does not mention the fact that Ajdukiewicz's discussion of the possibility of constructing languages without any axiomatic meaning rules and of the feasibility of the radical empiricist view of logic had nothing to do with Quine's philosophy but was prompted by the claims made earlier by Łukasiewicz (that experience may decide which of the alternative logics is the true one), von Neumann and Reichenbach (that quantum theory requires a non-classical logic), *etc.* Ajdukiewicz made it quite clear that in his view changes in language are distinguishable from changes in empirical claims and that one's logic may be altered either through a change of language (by altering its axiomatic rules) or through the joint testing of various assumptions (in a language without axiomatic meaning rules).

(3) I guess that it was editorially expedient to say in the Preface (p. xi) that Polish philosophers in the interwar period—and Ajdukiewicz was one of them—'like the Wittgenstein of the "Tractatus" . . . took philosophy to be "the logical clarification of thoughts" '. Historically, however, it would have been more correct to say that any such affinity between the author of the 'Tractatus' and Polish philosophers of the interwar period was due either to some common philosophical antecedents (*e.g.* Russell) or to the intellectual atmosphere prevailing at the time in Vienna and in Warsaw. For Wittgenstein's philosophy itself or his view of philosophy had no direct influence on Polish philosophy of the time at all. His 'Tractatus' did not command any respect and was largely ignored. As regards Ajdukiewicz, I can remember only one single (indirect) reference to Wittgenstein in his writings (in 'Meaning and Language'); and even that single reference was clearly intended to establish the independence (in the psychological and historical sense) of Ajdukiewicz's conception of empirical meaning rules from Wittgenstein's verifiability conception of meaningfulness. Ajdukiewicz even succeeded in writing an historical account—admittedly short but good—of the development of the philosophy of logical positivism[1] without once mentioning Wittgenstein's name.

(4) It is a pity that in their list of Ajdukiewicz's works (p. xiv) Skolimowski and Quinton did not mention his articles available in English. Though, for reasons unknown to me, there was no Polish contribution in Ayer's 'Logical Positivism', Ajdukiewicz's 'The scientific world-perspective'—one of the three

---

[1] Ajdukiewicz [1946], pp. 155–76.

or four pre-war articles on radical conventionalism—may be found in Feigl and Sellar's *Readings in Philosophical Analysis* [1949]. Perhaps the most important of the mentioned articles 'Language and meaning' and 'The Worldview and scientific apparatus' have been made available to English-speaking readers by John Wilkinson (in mimeographed edition of The University of Chicago) with very useful explanatory and critical annotations. A shortened version of 'Epistemology and semantics' can be found in *Proceedings of the Tenth Congress of Philosophy* (Amsterdam, 1948); 'Logic and experience' in *Synthese*, **8**, 1950; 'On the notion of existence' in *Studia Philosophica*, **4**, 1951; 'Three concepts of definition' in *Logique et Analyse*, 1958; 'The axiomatic systems from the methodological point of view' in *Studia Logica*, **9**, 1960, 'From the methodology of deductive sciences' in *Studia Logica*, 1966, 'Intensional expressions' and 'Proposition as the connotation of sentence' in *Studia Logica*, 1967.

(5) On p. xvii of the Preface we are told by Skolimowski and Quinton that '. . . in a rare departure from strict philosophical abstraction . . .' Ajdukiewicz also wrote an article on the freedom of science. This statement strongly suggests that Ajdukiewicz had very narrow interests and spent most of his life in the ivory tower of abstract philosophy; perhaps even, that he was cut off from the rest of the society after Schaff's ideological attack on him (reported on p. xiii). Nothing could be farther from the truth than such an image of Ajdukiewicz's personality and life. Admittedly the translators' Preface also mentions four years served by Ajdukiewicz as rector of Poznan University, before he moved to Warsaw in 1952, and the fact that he remained chief editor of *Studia Logica* until his death in 1963. But the important point is that throughout his whole life Ajdukiewicz took a keen interest in practical, moral and political issues (the mentioned article was, in fact, his second on the same topic, and both were preceded by articles on 'The value of science', 1938, 'Human action', 1938, 'On justice', 1939, 'In praise of active life', 1948) and spent a vast part of his time and enormous effort on teaching, reforming curricula (for secondary schools and universities), analysing methods of teaching (his pioneering work on the logic of questions and answers was primarily motivated by these considerations), organising research and organising regular symposia and conferences in which—apart from logicians and logically minded philosophers—also eminent scientists and mathematicians took part occasionally. His activities in this area in post-war time created for the logic based philosophy an exceptionally favourable atmosphere. Financially this was made possible by his links with the Polish Academy of Science. Though he never made any ideological concessions either in his writings or in his public appearances, he was rarely seriously obstructed in his activities. His works were never ignored and no serious scholar was ever afraid of making references to them. His retirement did not alter his pattern of life. Until his death he was popular and respected. In turn, he enjoyed his role and position. By contrast, he returned from his tour of the United States and of Britain (in the late 'fifties, I think) rather depressed and disappointed by factional squabbles among philosophers and by the erosion of the sense, so strong in his own generation, of participating in a worthwhile, universal philosophical enterprise.

JERZY GIEDYMIN
University of Sussex

REFERENCES

AJDUKIEWICZ, K. [1934]: 'Sprache und Sinn', *Erkenntnis*, 4, pp. 100–38.

AJDUKIEWICZ, K. [1946]: 'O tzw. neopozytywizmie', *Mysl Wspolczesna*, 6–7.

AJDUKIEWICZ, K. [1947]: 'Logika a doswiadczenie', *Przeglad Filozoficzny*, 43, pp. 3–21.

AJDUKIEWICZ, K. [1964]: 'Zagadnienie empiryzmu a koncepcja znaczenia', *Studia Filozoficzne*, 36, pp. 3–14.

POINCARÉ, H. [1902]: *Science and Hypothesis*.

POINCARÉ, H. [1905]: *The Value of Science*.

QUINE, W. [1969]: 'Epistemology naturalized', in *Ontological Relativity and Other Essays*.

BARTLEY, W. W. [1973]: *Wittgenstein*. London: Quartet Books. Cloth £2.75, Paper, £1.50. Pp. 145.

Bartley's extraordinarily concise new book on Wittgenstein raises and answers some new questions about its hero as well as answering other questions that spring to the reader's mind—both important and trivial. Even rumours are hunted down and settled: those of Jewish descent, flirtation with Roman Catholicism, election to the exclusive Cambridge University club The Apostles, homosexuality, are all true, apparently; those of giving away a fortune and of a scandal involving sadism, at most half-true. Bartley even looks into the interpretation of Wittgenstein's dreams; his concern for as well as his self-assertion towards both his poor Austrian schoolchildren and his privileged Cambridge students; his humility, vanity, inner struggles, conceit; as well as his devotion to others and capacity to dominate them. The book is more interesting and provocative than anything else written about Wittgenstein, whether intellectual commentary or personal recollection. It offers a vivid pen-portrait, which is full and satisfying. This is an astonishing feat from an author who is a philosopher rather than a biographer, and who never met his subject.

At the centre of Bartley's book are two problems collapsed into one. First, what brought about the transition between Wittgenstein's earlier and later philosophies? Second, what did Wittgenstein do in the 'missing' years between the two? Accepting the standard view that Wittgenstein produced two philosophies and that the later philosophy is to some extent a repudiation of the earlier, Bartley also takes over the more recent view of Engelmann that there is nevertheless a continuity of problems between the two. Bartley is thus involved at once in the controversies surrounding the interpretation both of the *Tractatus* and of the later works in order to set out just what the transition is that he is trying to explain, and how and where it took place. The odd thing is that Bartley in fact offers no intellectual explanation of the transition. Instead, he tells us what Wittgenstein did in the missing years. Thus he manages to collapse the (internal) problem of Wittgenstein's intellectual change into the (external) biographical problem of what happened to Wittgenstein after he had finished the *Tractatus* in August 1918 and before he resumed professional philosophising in Cambridge in January 1929. So far as I can follow it, Bartley tries to show how Wittgenstein's character and inner compulsions, his position in the Viennese aristocracy of wealth and culture, led to his undertaking practical and intellectual work (elementary schoolteaching and house design) in the course of which he refuted his own finished philosophy. He further thought he had undermined the very possibility of constructing a finished philosophy. Anything as cut and dried as a finished philosophy was *ipso facto* on the wrong track. We may ask whether the practical activity and the criticism were connected. Bartley seems to think they were. I cannot see where he establishes the point.

As to the interpretation of that finished philosophy, the *Tractatus*, Bartley argues negatively that: it is only incidentally a work in the foundations of mathematics and logic; it is not, as Quinton claims, about how language is possible; it is not a positivist tract against the possibility of metaphysics. Above all, Wittgenstein's finished philosophy is not a work that grows organically out of the tradition of British empiricism (*pace* Pears [1967]). Bartley is at pains to show how ironical it is that Wittgenstein was adopted by British philosophers as one of their own, when his roots, his outlook, and his philosophical concerns—his finished philosophy—were all those characteristic of the Vienna of the time, now extinct. But Bartley also rebuts the idea that the *Tractatus* and all that grew out of the Viennese concern, at the turn of the century, with a so-called crisis of language. What, then, on the positive side, is the book about? About ethics, politics, and aesthetics: topics about which the *Tractatus* says almost nothing, and what there is is terse and cryptic. This is all according to plan, because—says the Master—these subjects cannot be talked about, they can only be shown.

Whence this doctrine of showing? Wittgenstein was deeply interested in the paradoxes, but thought that his teacher Russell's theory of types was unnecessary to avoid them, since full understanding of the meaning of a logical or mathematical symbol contains knowledge of its range of applicability and hence full and proper understanding itself suffices to prevent us from getting into paradoxes. Similarly, a proper grasp of philosophical concepts is sufficient to prevent us getting into metaphysical conundrums, paradoxes, antinomies.

There are pointers, then, in the views of the early Wittgenstein, towards those of the later Wittgenstein: grasping the sense and reference of a symbol (concept) includes knowing how to apply (use) it. Thus does Bartley show us that the later views are remnants of the collapse of the earlier ones. What led to the collapse in the first place he does not say.

The other question is what happened to Wittgenstein after he had sent off the manuscript of his *Tractatus* on its round of publishers. Bartley describes Wittgenstein as the senior scion of a wealthy, philanthropic, and culturally entrenched Viennese family, tortured and depressed by the war and its aftermath. His country had been ravaged and impoverished, a close friend had been killed, suicide hung like a cloud over the family, and Ludwig himself was overwhelmed by intense homosexual urges which both obsessed and disgusted him. What could he do? What he did was to get a teaching qualification and exile himself from the horrors and the temptations by teaching school in remote and primitive peasant villages. Although not a devotee, Wittgenstein was in general agreement with the Austrian *Schülereform* movement of the time, which emphasised active, exploratory teaching, not learning dogma by rote. His methods were thus explosively new to the locals; his intellectually provocative way of employing them enthralled the students and enraged their parents.

During this time, Wittgenstein worked directly on the problem of language learning, and wrote a *Wörterbuch*—until Bartley's study, an ignored volume and a source of embarrassment to those familiar with it. It was then that he discovered meaningful communication did not rest on plain and simple atomic sentences or anything resembling them. So any very literal version of logical atomism—the doctrine he had himself adumbrated in the *Tractatus*—could not be true. Wittgenstein's *Wörterbuch* is largely a spelling dictionary, hence disappointingly unworthy of the Great Master in the eyes of disciples. Bartley, though not a disciple, finds a wealth of material in it. First, respect for dialect, *i.e.* for the common man and his common speech. Second, Wittgenstein shows there how much context and use affect the meaning of even simple or dialect words; words, then, must be defined against their context. Contexts, however, unlike definitions, are flexible, alterable, open-ended. It was the consequences of this new open-endedness which Wittgenstein had discovered for himself that he was to spend the rest of his life developing.

New to Wittgenstein as all this open-endedness was, just how original was it? This is a question into which Bartley does not explicitly enter, but on which there is a literature readers of this *Journal* may care to know about. Feyerabend, in a cryptic footnote to his classic and much-anthologised review ([1955]) denies originality and invokes the names of H. Gomperz, E. Mach, D'Alembert and K. Popper; Gellner [1959], pp. 215–7, finds most of it formulated and spurned in Hume; Gershon Weiler [1970] flatly claims 'Many of the criticisms levelled by Wittgenstein in his *Philosophical Investigations* against the views he had expressed in the *Tractatus Logico-Philosophicus* are to be found, even if less pointedly, in Mauthner' (p. 4.)

Although the present book is not primarily a critical work, just enough is said, and at just the right junctures, to indicate the author's view of his subject's philosophy: he seems to think it is almost totally misconceived. Although Wittgenstein achieves brilliant insights now and then and unquestionably has a most unusually powerful mind, suggests Bartley, his crucial ideas are both irrelevant and vague. Respectful as Bartley says he is, he demolishes Wittgenstein too thoroughly to make his expressions of respect convincing—at least without further explanation, which is not forthcoming. Bartley indicates that some of the central argumentative props of both the earlier and the later philosophy cannot stand. Wittgenstein's convoluted argument against his own earlier philosophy—that atomic sentences are not required for meaningful communication—lacks force. We all agree communication may not require atomic sentences, the question is whether its contradictory is an essential corollary of all or even some brands of logical atomism. Wittgenstein never bothers to show that this is so, and thus his self-refutation fails. Part of the trouble may be that Wittgenstein never clarifies the concept of atomic sentences, indeed he never gives an example of one. This vagueness at the centre of things recurs in the later philosophy, where the crucial and tantalising concept of form of life is not discussed, illustrated, defined, or clarified in any other way. Charging the apostle of clarity with so much vagueness is no small matter. All this, it seems, forces Bartley to explain Wittgenstein's popularity in the English-speaking world in some other way than by the power and interest of his ideas; in fact he does it by following Gellner's suggestion ([1959], p. 247–8) that they provide an open-ended (if misconceived) 'research programme' to anxious academics suffering thesis-pressure and publication-pressure.

The cumulative effect of all these various lines of approach is to cut Wittgenstein down to size, as it were. Wittgenstein's only claim to be included in Walter Kaufman's 'Portraits' series (other names: Sartre, Engels, Luther, Lenin, Lincoln) is that a lot of academic philosophers regard themselves as his followers, and the word of his alleged importance has even filtered out to the general reading public. They pick up Professor Bartley's book and read an absorbing —even mildly sensational—account of a famous philosopher. Do they also appreciate the ultimate irony that the seemingly very respectful book, though indeed respectful towards Wittgenstein the man, subtly debunks Wittgenstein's ideas and openly ridicules those who have become his followers and commentators? Wittgenstein is pen-portrayed as a brilliant, tormented figure, personally fascinating, but a minor-league thinker, constantly misunderstood, and almost sure to have been opposed to, if not plainly disgusted by, the secondary literature industry which he inadvertently sired. Indeed, Wittgenstein was the father of *two* research programmes and he certainly repudiated in no uncertain terms both that (logical positivism or the elimination of metaphysics through the logical analysis of language) which was claimed to grow from his *Tractatus*, as well as that part of the second he lived to see. There is something irresistibly tragicomic in Bartley's picture of armies of eager-beavers busily doing homage to Wittgenstein as best they know how by carrying on his programme. Some of them are going to pick up Bartley's book and get very angry when they find the joke is on them.

If the volume has a weakness, it is that it is several books rolled into one (biography, commentary, critique). Alternatively, one could say that it operates on several different levels of intellectual difficulty (narrative, interpretation, critique). Or again, that it is at the same time straightforward, a trifle slick, and yet profound. For all this, the book works. Bartley's ability to handle the impossible, to write enthrallingly about the philosophy as part of the whole man, whets the appetite as to who or what will be the subject of his next study. Another Austrian schoolteacher, perhaps?

I. C. JARVIE
York University, Toronto

REFERENCES

FEYERABEND, P. K. [1955]: 'Wittgenstein's Philosophical Investigations', *Philosophical Review*, 64, pp. 449–83.
GELLNER, E. A. [1959]: *Words and Things*.
PEARS, D. [1967]: *Bertrand Russell and the British Tradition in Philosophy*.
WEILER, G. [1970]: *Mauthner's Critique of Language*.

NELSON, L. [1970–1]: *Progress and Regress in Philosophy*. (Translated from the German by Humphrey Palmer.) Oxford: Basil Blackwell. 2 vols, Vol. I (with a preface by P. Bernays and an introduction by J. Kraft). Pp. xiii + 256. Vol. II (with a postscript by J. Kraft). Pp. xiii + 305.

This philosophical work of Leonard Nelson (1882–1927) is based on a course of lectures given at the University of Göttingen between 1919 and 1926. The material was assembled and edited posthumously by the late Julius Kraft, published for the first time in Germany in 1962 and now admirably translated into English by Humphrey Palmer. The title is an allusion to Jakob Friedrich Fries's work on the history of philosophy (1837) which, translated into English, reads: 'The History of Philosophy, written following the progressive steps of its scientific development'.

The book is divided into four main parts: pre-Kant, Kant, regress after Kant, progress after Kant (= Fries). I shall deal with the last part first, because this is the best way to understand the rest. I cannot really agree with Kraft, who in the Postscript praises Nelson for his 'clarity in the sequence of thought' (Vol. II, p. 287), or for his 'transparancy in its development and presentation'. One could have wished that Kraft had told us how the original material was organised and what criteria he followed in his editing.

Nelson's intentions as a teacher of history of philosophy (as expressed in the introductory 'General Plan of this work') are clear enough. Nelson stood in direct opposition to the tradition epitomised by Windelband's dictum: 'the more explicit and definite the standard of judgment is, the more partisan, unfair and useless as history the resulting narrative be' (Vol. I, p. 12). But while Nelson emphatically rejects this, the standards by which he judges philosophical ideas of the past are, for the most part, tucked away in the last 150 pages of this 600-page book. I hope in the present review to do for the reader what Nelson so signally failed to do: namely, to summarise as simply as possible the main tenets of his philosophy (including his philosophy of science). This task is complicated by the fact that Nelson's thinking is dominated by his re-discovery of Fries (1773–1843). For the most part, Nelson's way of presenting his own ideas is by expounding the philosophy of Fries. It is to this that I now turn. (So far as I have been able to check, Nelson's transcription of Fries is always faithful to the original.)

Nelson regarded Fries (whose first treatise appeared in 1798 when Fries was 25 and contained the essentials of his whole theory) as Kant's only worthy successor. Fries's first treatise was entitled *On the Relation of Empirical Psychology to Metaphysics*. This work displayed a thorough-going justificationism: every valid belief must be backed by 'positive' evidence, given unshakeable grounds, grounded on an enternally firm basis. And when I say 'every' I mean 'every'. According to Nelson, Fries accepted Kant's *Critique of Pure Reason* as a justification of the principles of metaphysics, the assumptions which make our scientific and indeed everyday knowledge possible; but, Fries asked, what justification is there for *Critique* itself? Fries rejected Kant's Transcendental Deduction. His aim was to provide a new kind of justification. In his very first paper, he presents a 'well-thought-out plan for a fundamental reform of logic, and especially of the logical theory about the justification of judgements' (Nelson, II, p. 165). Now I shall embark on an unavoidably long summary of this theory.

The key to Fries's logical theory is a distinctive conception of what a judgment is. It is something different than for Kant: 'Judgment depends on a combination, in itself quite arbitrary, of concepts, which we can make this way or that, as we please. . . . So the first question is always why a particular combination should be singled out. . . . Hence the need to justify each judgment, *i.e.* to derive each such assertion from some other item of knowledge. For a judgment becomes an item of knowledge only when it is based on some other item of knowledge, which the judgment *simply recapitulates*. The agreement of the judgment with this item of knowledge on which it is based is the test of its truth, and to demonstrate this agreement is what is called justifying the judgment' (II, p. 170).

This definition of judgment indicates the basic difference between Fries's and Kant's notion of a judgment: for Fries there is some knowledge which is *not* in the form of a judgment (II, p. 170). This is what he calls 'immediate knowledge'. For Fries, a judgment is a recapitulation (*Wiederholung*) of other items of knowledge, some of which are immediate and non-judgmental (or non-propositional). Fries's plan for the fundamental reform of logic, as faithfully reported by Nelson, adds up to the following: every judgment requires justification. If a judgment could be justified only by derivation from other judgments, there would be an infinite regress of justification. So there must be items of immediate knowledge which do not require justification 'if we are to be able to speak of truth at all' (II, p. 171). Nelson epitomises Fries's view thus: 'The question of justifying immediate knowledge does not arise. It ought not even to be asked' (II, p. 172). Here, the old Kantian epistemological optimism (we *do* have knowledge—how is this possible?) is re-named by Fries the 'Principle of the Self-Reliance of our Reason'. For Nelson (II, p. 172) as for Fries, it would be a mistake to regard the impossibility of justifying immediate knowledge as an admission of scepticism: 'Quite the opposite! Immediate knowledge requires no justification, for the simple reason that it already contains as immediate fact just what such a justification would be meant to gain for us. This fact must be accepted as such. It is not a problem for us, it is our situation. All we need is to grasp it and recognise it' (II, p. 173).

For Fries, the method of proof, *i.e.* logical derivation of judgments from other judgments, is only *one* stage of the process of justification. It leads back to so-called basic judgments which are justified in their turn by immediate knowledge (II, p. 175). So what is immediate knowledge? There is both perceptual and non-perceptual immediate knowledge, but whether it comes from the outside (perception) or from the inside (metaphysical immediate knowledge), it is not learnt: it comes as an already completed datum.

It is actually non-perceptual immediate knowledge that is the key to the philosophy of Fries. (Those who have met his name only in Popper's *The Logic of Scientific Discovery*, are liable to assume that immediate knowledge is just perceptual knowledge.)

Before saying more about non-perceptual immediate knowledge, let me finish the exposition of Fries's theory of method. It is clear that in order to avoid the infinite regress there must be a method of justification beyond what he called 'proof', *viz.* derivations of statements from statements. This method, as I have already said, consists in tracing the basic judgments (*viz.* the premisses of a 'proof') to immediate knowledge. If the immediate knowledge is perceptual, the justification is called a 'demonstration'; if non-perceptual, a 'deduction'. Nelson emphasises (II, p. 177) that for Fries, as well as immediate perceptual knowledge, we also have immediate consciousness of our perceptual knowledge. This latter makes superfluous any attempt to justify the former.

The other and more original method of justification is ***deduction*** which applies to the basic judgments of metaphysics. 'Basic judgments of metaphysics', Nelson explains, 'are those about God or the world, about causality or substance, about duty, goodness, freedom and beauty' (II, p. 181). These judgments are traced back *to immediate metaphysical knowledge.* Where does this knowledge come from? Fries postulates *three* sources of knowledge: reflective intellect, sense, and thirdly (in Nelson's words) 'a pure reason independent both of sense and of intellect, an autonomous faculty of knowing which is native to the mind and a permanent characteristic of it, and which unites all individual items of knowledge into a whole' (II, p. 207).

According to Fries, the failure to recognise this third source of knowledge has been the main error of traditional philosophy. Kant, with his three types of judgment, should have freed himself from the 'dogma' of only two *sources* of knowledge, reason and experience. The discovery of this new *source* of knowledge was, for Nelson, the major step forward in philosophy after Kant.

How can we get at this source? We have no direct consciousness of metaphysical truths; so we have to undertake a 'detour'; this consists in an empirical, psychological enquiry 'into immediate metaphysical knowledge, to make sure that it does exist, and what it contains—thus acquiring a knowledge of it which is distinct from the judgments of metaphysics and which allows us to *justify* them (this is what is called deduction)' (II, p. 181).

According to both Fries and Nelson, 'deduction' is not a *proof* of the metaphysical judgments. The proof which Kant sought to provide through his Transcendental (Nelson calls it 'Objective') Deduction of the Categories is rejected by them: 'No proof . . . of the principles of metaphysics is possible at all' (I, p. 200). Fries explicitly says that philosophy should abandon the prejudice that all truths are provable. 'Deduction' consists in 'revealing' (*aufweisen* as against *beweisen* (= prove)) the original knowledge, hidden in our reason, which 'underlies' (*zugrunde liegen*) metaphysical judgments. It is through 'induction from inner experience' (II, p. 209) that Fries tries to establish a Theory of Reason, a kind of 'Natural Theory of the Human Soul', an 'inner experimental Physics'. 'The Critique has indeed to justify the principles of metaphysics, but that does not mean it should prove them, only that it should reveal (point out) their epistemological grounds' (I, p. 188). 'I do not prove', Fries says in the *Neue oder Anthropologische Kritik*, 'that all substance is permanent, rather I *point* to the fact that the principle of the permanence of substance lies in every finite mind'.

Nelson gives the (not very happy) name 'Speculative basic form' to that which is *obscurely* present, but capable of illumination, in human reason; and he says: 'the speculative basic form is what makes categories possible. Metaphysics therefore possesses in this speculative basic form an unassailable foundation like that possessed by geometry in the perception of space' (II, pp. 208–9).

From this quotation we see that Categories are not innate; what is innate is the 'speculative basic form'; and this is open to empirical, anthropological investigation. It is present in every human mind. Our knowledge of the categories is gained through reflection which does no more than 'recapitulate' what was already in us obscurely and bring it into the clear light of consciousness.

In this way Nelson hoped to resolve Kant's dichotomy of *quaestio facti* and *quaestio iuris.* He used a purely psychological procedure to show the existence of something prior to all metaphysical judgments. And this provides him with the epistemic ground (*Erkenntnisgrund*) he was searching for. According to him, we have a situation here analogous to that with empirical judgments which have their 'ground' in immediate perceptual knowledge. This 'ground' is their source. For both Nelson and Fries, the *quaestio facti* and the *quaestio iuris* are solved together by the process of 'pointing out' such 'grounds'.

Nelson now claims that whereas any attempt to *prove* a metaphysical or an empirical judgment (*i.e.* to derive it from prior judgments) would involve an infinite regress, no infinite regress is involved in either a 'demonstration' or a 'deduction'. For the process of 'pointing out' does not itself figure as a premiss in the subsequent 'deduction'.

This brings to an end my reconstruction of Nelson's account of Fries's Theory of Method. If anybody has succeeded in following me so far, he will share with me the enjoyment of smiling over Fries's statement that 'this simple theory of the justification of judgment was enough to make philosophy as simple a school subject as geometry and arithmetic had been for thousands of years' (II, p. 190). Nelson's presentation of it suffers in being diluted over 600 pages. I would suggest to anybody seriously interested in Fries's philosophy who can read German, but has not the time to read his original works, to look up Chapter 7 of volume 3 of Cassirer's *Das Erkenntnisproblem* before tackling Nelson's book. Cassirer's exposition, including some good criticism, is much more succinct (30 pages) and easier to understand.

Now I want to probe into Nelson's theory of philosophical progress and try to uncover some of its implicit presuppositions.

Two basic assumptions led Nelson to 'posit' (II, p. 202) a non-perceptual immediate knowledge: (1) metaphysical judgments can have 'objective' validity; (2) but this would not be possible 'unless one can point out the basis on which they could be known' (II, p. 215). He holds that judgments can be valid though unprovable. He rejects the 'epistemological axiom that every truth must be open to proof; an axiom that is certainly false, since if it were true its contrary would follow from its own unprovability' (II, p. 47). However, he also holds that every true judgment must be open to justification. This is the second assumption. (He appears not to have asked whether *this* is self-refuting.) With Nelson, rationalism *is* justificationism.

Fries and Nelson had no clear conception of induction and did not try to work out a theory for it. In consequence Fries (and Nelson after him) had no right to claim universal validity for the results of his psychological-empirical Critique. (See Fries's *Neue Kritik Der Vernunft*, p. xx, where he speaks of the 'infallible' critical method.) Since Nelson himself says (II, p. 119), 'We can never assert the validity of these (empirical) laws with unqualified rigour, but only that experience has so far vouched for them', he should have conceded that Fries's method yielded contingent results, open to revision. But this is not the impression one gets from Nelson, *in spite* of all his insistence that metaphysical principles cannot be proved (*bewiesen*) but only pointed out (*aufgewiesen*).

One cannot understand many of Nelson's remarks on Descartes, Locke, *etc.*, unless one realises that he took over Fries's view of truth, which Nelson introduces whilst criticising Kant only halfway through the first volume. Fries rejected what we now call the semantic conception of truth (correspondence theory of truth). In Nelson's words, 'The relationship of knowledge to its object is not a topic for scientific investigation at all, for it is impossible to compare knowledge with its object' (I, p. 191). For both men, truth is a relationship between items *within* our knowledge. (Nelson on several occasions suggests that the expression 'it is true' is redundant.) Fries's conception of truth was bound up with his justificationist theory of knowledge. His starting-point was the need to escape the infinite regress of justification. He hoped to halt the regress at 'immediate knowledge'. This he called 'rational truth'. We have 'intellectual truth' when a judgment (*i.e.* immediate knowledge) is in agreement with immediate knowledge.

But, of course, the regress cannot really be halted by a 'recapitulation' of an unutterable something. (Fries tried in vain to overcome the problem of unutterability of immediate metaphysical knowledge, by calling it 'obscure': it can be made conscious 'only step by step'.) Such a 'recapitulation' cannot really yield the remarkable bonuses which Fries and Nelson claim: not only the Categories, but various higher level, universal theories, like the Constancy of Mass, the law of inertia and the equality of Action and Reaction.

Fries rejected what is now known as the doctrine of the unity of science. In Nelson's words, Fries enunciated the principle 'that truth splits up into different world-views' (II, p. 248). Fries argued, or asserted, that there cannot be any systematic unity of human knowledge: 'the various modes of judging the existence of things go their separate ways', they cannot be brought together under common principles. The world is one, but our knowledge of it is split up. We have a mechanistic (theoretical), a morphological (common sense), a psychological and an ethico-political world-view. There is also a religio-aesthetic world-view which is said to be formed 'by negating the bounds' of all the naturalistic world-views (II, p. 277).

The mechanistic world-view, according to Nelson, is the metaphysics of external Nature based on the only natural phenomenon that can be constructed mathematically, namely motion; it was worked out systematically by Fries in his book *Mathematical Philosophy of Nature*, a book 'which got buried under the rubbish of the so-called speculative philosophy of nature of the school of Schelling and Hegel' (II, p. 251).

But Nelson, unlike Fries, was faced by the new development of relativistic physics. His reaction was subdued, almost defeatist: 'For the present it does seem as though natural philosophy and empirical physics will have to go their separate ways' (II, p. 253). Moreover, it would be hasty to draw metaphysical consequences from a theory (modern physics) which is in such a muddle that we 'are not even in sight of satisfying the logical requirement of non-contradiction' (II, p. 253). He now says that the only position which can be reconciled with a critical philosophy of nature is the 'self-denying stance of Conventionalism', a position he had sharply criticised (I, p. 59). Nelson openly admits that Einstein's critique of the concept of simultaneity, if taken in a non-conventionalist way, would force us to reject Kant's third analogy of Experience, the principle of reciprocal action which is equivalent to the assumption of action at a distance (II, p. 255). But at a realistic, metaphysical level, Nelson refuses to reject action at a distance; for that would involve a reversion to Cartesianism, to action by contact, a metaphysical theory which for Nelson cannot be true. If some future generation tries to build a bridge between Kant's metaphysics of Nature and 'experimental physics', it must infallibly revert to some form or other of the postulate of Mechanism' (II, p. 256). He does not stand completely alone among philosophers of the period in his metaphysical opposition to relativity theory. (Hugo Dingler, among others, also opposed it in his *Physik und Hypothese* [1921].) Nelson reveals here a fair intuition in scientific matters when he insists on regarding the Michelson–Morley experiment, not as the 'basis' of Einstein's discovery, but as a 'contingent occasion for the relativistic critique of the old spatio-temporal conception' (II, p. 254).

It is only after we have become acquainted with Fries's philosophy in the last quarter of Nelson's book that the first three-quarters of the book become exciting and interesting to read. These deal with the history of philosophy. At first reading they are very puzzling. One needs to understand the meaning of every phrase because he is extremely succinct. (Descartes, Locke and Berkeley receive a total of twenty pages.) Without a grounding in Fries's theory of method and justification, the reader (at least *this* reader) cannot appreciate why, for example, Nelson calls Descartes' attribution of error to human wilfulness a 'psychological discovery' (I, p. 32) which was a 'permanent gain for epistemology'. For the reader does not yet know of Fries's thesis that error is possible only in non-immediate knowledge and that the faculty of reflection involves arbitrary choice. Nor does he know about the *psychological* nature of Fries's Critique.

Descartes, for Nelson, is rightly called the Father of modern philosophy because both the rationalists and the empiricists adopted 'the dogmas he had laid down as axioms'; not even Kant succeeded in completely freeing philosophy from them; only Fries did. Some of these dogmas were older than Descartes; for example the dogma that all knowledge comes either from the intellect or from the senses, which goes back to Aristotle's dichotomy of logic and experience. Now we are able to understand why Nelson insists so much that this is an unwarranted assumption on which no true philosophy could be built: it does not leave room for the (by now notorious) 'metaphysical, non-perceptual immediate knowledge'. A reader who has not read this review and was not whimsical enough to read the book backwards, cannot understand why Descartes is so wrong in regarding only inner, but not outer, perception as immediate knowledge.

Nelson's style is in these early pages fast-moving. He summarises the problems and sharply criticises the solutions offered by these philosophers in a few sentences. His criticisms are usually aimed at their method of arguing, their 'dialectic'. He pounces on any logical sleight of hand or surreptitious shift of meaning. Arguments based on the ambiguity of a particular word are analysed and rejected. This task is performed with almost the zeal of a linguistic philosopher; but he is always hammering away at real problems. These few pages contain an astonishing amount of philosophical information. Two of the best chapters are those preceding and immediately following the twenty pages on history of philosophy proper. The title of the former is 'The Dispute about Universals, and the Rise of Natural Science'. The problem of universals was solved by natural science, says Nelson (I, p. 23), which does not worry about concepts but concentrates on theories and laws. The title of the other chapter is 'Excursus on the Physics of Galileo and Newton and its Philosophical Significance'. Here he attacks conventionalists for not recognising the decisive role of metaphysics in science, and accuses conventionalists of misusing nominal definitions ('This is where the cheating always starts') (I, p. 60). We meet here the familiar names of Boltzmann, Mach, Plank, Poincaré, le Roy, Osiander Bellarmino. It is in this chapter that one of Nelson's 'most far-reaching assertions', as Julius Kraft calls it in the *Postscript*, makes its first appearance: the thesis that strict empiricism is not better than idealism as a philosophy of science.

He does not devote a separate chapter to Leibniz and Spinoza, but the book abounds with references to them, which indicate with great clarity their basic thoughts. It is with Hume that Nelson grows really excited and cannot any longer keep his own philosophy in the background. His exposition of Hume's theory is followed by a chapter on its 'defects' and a second one on his 'lasting results'. At this point the reader begins to feel lost, because while some of Nelson's points are good and clear, the relevance of others cannot be fully appreciated until one becomes acquainted *in detail* with Fries's philosophy. An example: Nelson calls Hume's reduction of all reflective ideas to impressions a step forward in philosophy. We shall find out much later that Hume's main contribution to progress in philosophy was his discovery that even everyday empirical knowledge depends on tacit metaphysical assumptions which cannot be derived from either sense-impression or logic. So the setting is ready for Kant to discover the synthetic *a priori*, and with him we come to Fries's teacher.

The exposition of Kant's philosophy, which takes 150 pages, begins with a masterful reconstruction of Kant's thought as contained in his pre-critical writings. This shows how Kant gradually freed himself from the metaphysics of Leibniz and Wolff.

The *Critique of Pure Reason* itself is expounded in an interesting way, but Nelson underplays the role of the Antinomies. Nelson is happy to include Kant's Refutation of Idealism, because it confirms his own thesis that Idealism is an unavoidable consequence of radical Empiricism (I, p. 161). The *Critiques* of Practical Reason and of Judgement are briefly dealt with, together with the *Metaphysical First Principles of Natural Science*. Nelson's first volume ends with 60 pages on Defects in Kant's Theory, some of which I touched on while trying to piece together the philosophy of Fries. One can hardly understand these pages without having read the chapters on Fries. Many of Nelson's disagreements with Kant, such as that over the *quaestio iuris*, are too far-reaching to be put forward in the form of piecemeal criticisms of Kant.

In the first half of the second volume we come to that part of the history of philosophy which, Nelson effectively shows, is a regress from Kant's philosophy. Nelson's interest in Reinhold, Fichte, Schelling, Hegel and Benecke is not merely historical: he is interested in their philosophies because they develop potentially dangerous theses in Kant's *Critique*, drawing out their more absurd consequences and thus making their falsity easier to detect. This is why Nelson takes so much trouble in criticising philosophers like Fichte and Hegel, whom he despises. It is excellent to find somebody of Nelson's calibre who actually reads and analyses the sentences of Fichte, Schelling and Hegel, explains the trivial logical fallacies in their reasoning, goes through their examples and 'proofs' step by step showing why they do not prove what they pretend to prove. Nelson is here at his best, especially in his treatment of Hegel.

An 'excursus' on Historical Materialism is also excellent. Nelson notices that some phrases in this theory, like 'in the long run', 'remove the assertion in question from the control of experience as to its truth or falsity' (II, p. 109). When discussing the theory that contradictions have a reality in material existence Nelson blows his top, using the most vigorous language (II, p. 111) of anger and disgust.

At this point, where Nelson's human side shows itself so strongly, I come to the end of this review. He was deeply concerned with the problems of his age. He saw the First World War as the result of 'some still undiscovered flaw in the foundations' (II, p. 4) of our culture, and he tried to find and destroy these flaws in so far as they were embedded in Hegelian philosophy. But when he moved to the more positive task of finding right foundations, he fared less well. If we say that Fries bequeathed a programme of empirical, psychological, anthropological investigation into immediate non-perceptual, metaphysical, knowledge, we must add that Nelson, in this book, does not carry the programme any further: for instance, he did not investigate whether intuitions of time are culture-invariant or whether the Hopi, say, have intuitions different from ours. (Why then is Nelson translated into English but not yet Fries?)

Fries's philosophy represents the most ambitious form of justificationism since Kant. It is based on the dogma that we *do* have true knowledge. (The existence of science is taken by Nelson as a fact, as unproblematic—the old Kantian optimism.) However, its surpeme justifying principles remain peculiarly irrelevant to real problems in science (and also, incidentally, in politics).

The translation is very good, very few mistakes, which are at II, p. 185, where the axiom of Parallels is completely mis-translated and the words 'to a calculus' are missing from line 5 of I, p. 58. In the list of German words I miss the word *aufweisen*, *i.e.* 'pointing-out'.

Nelson uses diagrams to illustrate his ideas. Some readers might find them useful, I did not particularly.

ANDREA MANNU*
London School of Economics

---

* [*Editors' note*] Since this review was received we have learned with deep regret of Andrea Mannu's death.

PROGRAMME OF MEETINGS

JUNE 1974

The Third Annual Conference of the Society for Exact Philosophy was held Thursday–Saturday, May 9–11, 1974, at the University of Alberta, in Edmonton, Alberta, Canada. Scheduled speakers were: Nuel D. Belnap, Jr. (University of Pittsburgh), Hector-Neri Castañeda (Indiana University), Dorothy Grover (University of Illinois, Chicago Circle), Marsha Hanen (University of Calgary), Clifford W. Hooker (University of Western Ontario), David Kaplan (University of California, Los Angeles), Karel Lambert (University of California, Irvine), Hugues Leblanc (Temple University), Charles G. Morgan (University of Alberta), John Woods (University of Victoria).

Membership in the Society is open to anyone. Annual dues are $2.00 (U.S.) and should be sent to Ronald Giere, History and Philosophy of Science, Indiana University, Bloomington, Indiana 47401.

ALAN ROSS ANDERSON MEMORIAL FUND

Alan Anderson died last December 5th of cancer, at the age of 48. As a memorial to his manifold contributions (both academic and personal), there has been established at the University of Pittsburgh an Alan Ross Anderson Memorial Fund whose initial purpose is to fund a prize or fellowship for a graduate student or junior faculty member in philosophy. Readers of this journal who wish to remember Alan Anderson in this way should send a cheque to the Alan Ross Anderson Fund, in care of the Department of Philosophy, University of Pittsburgh, Pittsburgh, PA 15260. U.S.A.

(1974 - 68th Year)

SCIENTIA

INTERNATIONAL REVIEW OF SCIENTIFIC SYNTHESIS

Scientific Committee

G. ABETTI - G. B. BONINO - P. CALDIROLA - L. CALIFANO - R. CANESTRARI - B. DE FINETTI - B. MIGLIORINI - G. MONTALENTI - E. PADOA - G. SANSONE - A. SESTINI - V. SOMENZI - S. TONZIG - L. TREVISAN

Editors N. BONETTI . A. DE MURTAS

"SCIENTIA" publishes original articles by the most eminent scientists and philosophers in all fields of knowledge.

Scientia is a synthetical review of the various disciplines. It offers full scientific information and gives an account of the problems and latest developments in research.

It is issued every two months, each number of about 200 pages. The articles are published in the original languages of the authors (Italian, English, French, German, Spanish). Each number contains a supplement giving the complete English translation of all the non-English articles. The number also contains: Reviews, Bibliographical notices, Reports.

Annual subscription: Italy L. 18,000. Abroad L. 24,000.

One number of the current year: L. 4,000.

One sample back number: L. 500 in stamps.

"SCIENTIA" - Via Vivaio, 16 - 20122 Milano

THE PHILOSOPHER'S INDEX

An International Quarterly Index

To Philosophical Periodicals

The **Philosopher's Index** is an up-to-date quarterly index of articles from more than two hundred and thirty-five major philosophy journals and related interdisciplinary publications. Articles are indexed by subject and author. Abstracts of the articles are published in each issue and are written by the authors of the articles. In addition, the Index contains a Book Review Index and a section entitled Translations in Progress.

The Quarterly, Paperback, $24 (Individuals: $12)
Annual Cumulative Edition, Hardbound, $30 (Individuals: $15)

U.S. and Canada:

THE PHILOSOPHER'S INDEX
Bowling Green University
Bowling Green, Ohio 43403 USA

Outside the U.S. and Canada:
D. Reidel Publishing Company/Dordrecht, Holland

The Journal of Philosophy

Subscriptions at $10.00/year; $7.00 to students
$12.00 to libraries and institutions

The trustees and editors of the *Journal of Philosophy* announce an increase in subscription prices, effective in January of 1974. Individual subscriptions will be $10.00/year; library subscriptions will be $12.00/year; student subscriptions will remain as they are: $7.00/year. An additional $1.00/year is charged for postage, to subscribers outside the United States, Canada, and Mexico.

Back issues, singly or in volumes, from the founding of the *Journal* in 1904, are available from this office.

720 PHILOSOPHY HALL, COLUMBIA UNIVERSITY, NYC 10027

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

VOLUME 12/NUMBER 1 JANUARY 1975

CONTENTS

I PATRICK DAY: *The Uniformity of Nature*
II MICHAEL A. SLOTE: *Existentialism and the Fear of Dying*
III DAVID H. SANFORD: *Borderline Logic*
IV ALWYNNE MACKIE: *The Structure of Aesthetically Interesting Metaphors*
V JOHN L. POLLOCK: *Four Kinds of Conditionals*
VI WILLIAM A. CHRISTIAN: *Domains of Truth*
VII BARUCH BRODY: *The Reduction of Teleological Sciences*
VIII STEPHEN NORRIS: *The Intelligibility of Practical Reasoning*
IX HAROLD I. BROWN: *Paradigmatic Propositions*

Annual Subscription: $18.00 to Institutions (*ex*cluding Annual Monograph, price $7.00)
$8.00 to Individuals (*in*cluding Annual Monograph)

Published by BASIL BLACKWELL
Orders to: 108 Cowley Road, Oxford OX4 1JF, England

THE
REVIEW OF METAPHYSICS
A PHILOSOPHICAL QUARTERLY
Edited by JUDE P. DOUGHERTY

Vol. XXVII, No. 4 June 1974 Issue No. 108

ARTICLES
Ernan McMullin *Two Faces of Science*
Karsten Harries *Hegel on the Future of Art*
R. E. Allen *Unity and Infinity: Parmenides 142b–145a*
John Fisher and Jeffrey Maitland *The Subjectivist Turn in Aesthetics: A Critical Analysis of Kant's Theory of Appreciation*
Terence Irwin *Recollection and Plato's Moral Theory*
DISCUSSION
Michael J. Loux *Kinds and the Dilemma of Individuation*
BOOKS RECEIVED
William A. Frank and Staff *Summaries and Comments*
PHILOSOPHICAL ABSTRACTS
ANNOUNCEMENTS
INDEX TO VOLUME XXVII

Individual Subscription $9.00 Student Subscription $5.00
Institutional Subscription $15.00
THE CATHOLIC UNIVERSITY OF AMERICA, WASHINGTON, D.C. 20017
Twenty-Year Cumulative Index, 1947-1967 is available for $3.00 a copy.

THEORIA
A Swedish Journal of Philosophy
Edited by GÖRAN HERMERÉN *and* KRISTER SEGERBERG

CONTENTS
VOLUME XL 1974 Part 1 (60 pp.)

Sten Sparre-Nilson: Two additional remarks on the logic of historical explanation
Peter van Inwagen: A formal approach to the problem of free will and determinism
Kit Fine: An incomplete logic containing S4
S. K. Thomason: An incompleteness theorem in modal logic
Reviews by R. Eugene Bales and Thomas E. Hill, jr.

THEORIA, which invites contributions in all fields of analytic philosophy, is published three times a year. Subscriptions and requests for back numbers should be sent to the publishers, Messrs. C. W. K. Gleerup, Box 1205, 221 05 Lund 1, Sweden.
Annual subscription is Skr. 35:—(approx. 7 US $).

MIND

A Quarterly Review of Philosophy

edited by Professor **DAVID HAMLYN**

Vol. LXXXIII JULY 1974 No. 331

Articles	
Expressions of Intentions	D. GUSTAFSON
Describing, Evaluating, and Moral Conclusions	C. M. CHERRY
Retained Knowledge	A. HOLLAND
Backwards Causation and Continuing	S. WATERLOW
On Our Knowledge of Matters of Fact	E. SOSA
Criteria for Plausible Arguments	J. AGASSI
Discussions	
Grice and MacKay on Meaning	R. T. GARNER
On "Levels of Rules and Hart's Concept of Law"	T. M. BENDITT
Negative Utilitarianism	A. D. M. WALKER
The Explanatory Power of Chomsky's Transformational Generative Grammar	J. SINGLETON
Freedom as the Non-Restriction of Options	W. A. PARENT
Freedom as the Non-Restriction of Options: A Rejoinder	S. I. BENN and W. L. WEINSTEIN
Nullity and Sanction	P. MULLOCK
Tractatus 5.542	M. COHEN
Rights Without Justice	T. D. CAMPBELL
Book Reviews	
Books Received	

Annual Subscription £2.50, single issues 80p

Published for MIND ASSOCIATION by

BASIL BLACKWELL 49 BROAD STREET OXFORD

Studies in History and Philosophy of Science

Studies in History and Philosophy of Science is a quarterly journal which analyses the genesis, evolution and logic of scientific thought. It attempts to unite historical and philosophical approaches to science.

Recent papers include:
Maxwell's Methodology and his application of it to Electromagnetism, A. F. Chalmers. **The Astronomy of Eudoxus: Geometry or Physics?** Larry Wright. **Conceptual Structures and Scientific Change,** Garry Gutting. **Whewell's Theory of Scientific Language,** Morton L. Schagrin. **Speculation in Physics: The theory and practice of Naturphilosophie,** Barry Gower. **Feyerabend and Galileo: the interaction of theories, and the reinterpretation of experience,** Peter K. Machamer. Edited by Gerd Buchdahl, Reader in History and Philosophy of Science, University of Cambridge and Professor L. L. Laudan, Chairman, History and Philosophy of Science, University of Pittsburgh. Published by Macmillan Journals Limited.

Annual Subscription £8.50 (£9.50 U.S.A. and Canada). Prices applicable only to orders starting before 31st December, 1974. Payment may be made in any currency at the current exchange rate. Orders must be accompanied by remittance. Cheques should be made payable to: Macmillan Journals Limited, Subscription Department, Brunel Road, Basingstoke, Hampshire RG21 2XS, England.

Zeitschrift für allgemeine Wissenschaftstheorie

Journal for General Philosophy of Science

Edited by ALWIN DIEMER . LUTZ GELDSETZER . GERT KÖNIG

Per year 2 issues of 208 pp. each Annual subscription DM 64,–

From the contents of vol. VI.

NACHRUF

auf Aron Gurwitsch. Aron Gurwitsch als phänomenologischer Wissenschaftstheoretiker, von Lester Embree

AUFSÄTZE

Aant Elzinga, Some remarks on a theory of research in the work of Aristotle

Dietmar Kamper, Hermeneutik—Theorie einer Praxis?

Heinz Kimmerle, Die Funktion der Hermeneutik in den positiven Wissenschaften

Hans Krämer, Zur Ortsbestimmung de historischen Wissenschaften

Nelly Tsouyopoulos, Die induktive Methode und das Induktionsproblem in der griechischen Philosophie

DISKUSSION

Patrick A. Heelan, Hermeneutics of Experimental Science in the Context of the Life-World

Theodore Kisiel, Commentary on Patrick A. Heelan's Hermeneutics of Experimental Science in the Context of the Life-World

Patrick A. Heelan, Comments on Professor Theodore Kisiel's Commentary on my Paper: Hermeneutics of Experimental Science in the Context of the Life-World

BERICHT

Theodore Kisiel with Galen Johnson, New Philosophies of Science in the USA. A Selective Survey

REZENSIONEN

L. Geldsetzer, Die Philosophie der Philosophiegeschichte im 19. Jahrhundert. Ein Beitrag zur Wissenschaftstheorie der Philosophiegeschichtsschreibung und -betrachtung (Studien zur Wissenschaftstheorie, Band III), Meisenheim 1968 (Lucien Braun, Straßburg)

R. Bubner, K. Cramer und R. Wiehl (Hrsg.), Hermeneutik und Dialektik. Festschrift für H.-G. Gadamer, 2 Bde., Tübingen 1970 (L. Geldsetzer)

A. Hermann u. a., Lexikon der Geschichte der Physik A—Z, Köln 1972 (Gert König)

ZEITSCHRIFTENSCHAU

Chr. Strosetzki, Zeitschriftenschau 1973

FRANZ STEINER VERLAG GMBH . WIESBADEN, GERMANY

Noûs

Nihil philosophici a nobis alienum putamus

A Quarterly Journal of Philosophy

edited by Hector-Neri Castañeda

NOÛS publishes quality essays and brief discussions on philosophical problems regardless of the author's philosophical school or point of view. From time to time, NOÛS will publish symposia, surveys of work recently done on specially selected philosophical topics, critical studies of recent books (rather than reviews), and catalogues of recent publications in philosophy. NOUS publishes the symposium papers of the A.P.A. Western Division meetings in its March issue.

A quarterly, NOÛS is published in March, May, September, and November each year. For *individuals* the subscription rates are: one year, $9.00; two years, $17.00; three years, $25.00. For *bona fide students:* $8.00 per year. For *institutions:* $16.00 per year (plus 10 per cent for agency subscriptions). Add 50 cents per year for all non-U.S. subscriptions. Make checks payable to INDIANA UNIVERSITY—NOUS, or to INDIANA UNIVERSITY FOUNDATION—NOUS for 2- or 3 year subscriptions.

VOLUME VIII NUMBER 2 MAY 1974

R. B. ANGELL—*The Geometry of Visibles*

TRYG A. AGER, JERROLD L. ARONSON and ROBERT WEINGARD—*Are Bridge Laws Really Necessary?*

ALONZO CHURCH—*Outline of a Revised Formulation of the Logic of Sense and Denotation* (Part II)

RICHARD GRANDY—*Some Remarks about Logical Form*

M. G. YOES, JR.—*Intensional Language and Ordinary Logic*

SCOTT WEINSTEIN—*Truth and Demonstratives*

FRED FELDMAN—*On the Extensional Equivalence of Simple and General Utilitarianism*

LAWRENCE ROBERTS—*Libertarianism and Statistical Laws*

Corrigenda

Recent Publications in Philosophy

Order from

The Secretary, NOUS, Department of Philosophy
Indiana University, Bloomington, IN 47401

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY

Edited by MARVIN FARBER

JUNE, 1974, VOLUME XXXIV, No. 4

CONTENTS

I. ARTICLES
Perception and the "Concrete Abstractness" of Science, by CHARLES HARTSHORNE
Why Should I Believe in Science?, by ERVIN LASZLO
Philosophical Implications of Unity in the Contemporary Arts and Sciences, by MENDEL SACHS
Intrinsic Value and the Intrinsic Valuer, by ROBERT EDGAR CARTER
How Well Can One Get to Know a Strawsonian Person?, by DONALD SIEVERT
Something Really Unheard Of, by VIRGIL C. ALDRICH
Contra Wittgenstein, by RICHARD H. SCHLAGEL
A Phenomenological Interpretation of Kant's Apriori of Taste, by DONALD B. KUSPIT
II. DISCUSSIONS, REVIEWS
III. NOTES AND NEWS, RECENT PUBLICATIONS

Subscription prices $8.00 per year for individuals; $10.00 per year for institutions. Separate issues, $2.00 and $2.50 respectively. Postage and handling for back issues only. Papers submitted for publication will not be returned unless accompanied by a self-addressed stamped envelope, or equivalent postage.
A Circular listing the main contents of the Journal since 1940 will be sent upon request.

Address all communications to

PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH
State University of New York at Buffalo, Buffalo, New York, 14226

INQUIRY An Interdisciplinary Journal of Philosophy and the Social Sciences
Edited by Arne Naess
and Alastair Hannay

Number 3 Autumn 1974 Volume 17

Objectivism and the Study of Man (Part II)	Hans Skjervheim
Impossible Objects	Karel Lambert
Belief and the Limits of Irrationality	Keith Graham
Discussions:	
Von Wright on Historical Causation	Elazar Weinryb
Selfishness and Capitalism	Tibor R. Machan
Review Discussions:	
Criminal Insanity (H. Fingarette *The Meaning of Criminal Insanity*)	Herbert Morris
Quinton's Half-Hearted Ontology (A. Quinton *The Nature of Things*)	T. L. S. Sprigge
Cosmic Biology and the 'Great Shift' (S. Langer, *Mind: An Essay in Human Feeling*, Vol. II)	Eva Schaper

Published quarterly by Universitetsforlaget, P.O. Box 307
Blindern, Oslo 3, Norway

Annual Subscription: Institutions – N. kr. 75.00, £4.90, $12.50
Individuals – N. kr. 60.00, £3.90, $10.00
Postage included

Revue Internationale de Philosophie

Direction Scientifique: Professeurs Ph. DEVAUX, de l'Académie royale de Belgique
H. GOUHIER, de l'Institut de France
Ch. PERELMAN, de l'Académie royale de Belgique
Direction Administrative-Rédaction: Mme Simone SERVAIS
99 avenue de l'Université, 1050 Bruxelles-Belgique
Tel. 02/48 03 70

La Revue Internationale de Philosophie consacre, depuis 1938, chacun de ses 4 numéros annuels à l'étude d'un philosophe ou d'un mouvement philosophique.

No. 106—1973—Fascicule 4

Sommaire

Georg Lukacs *Tactique et Éthique.* Traduction de R. Legros & G. Haarscher

Gyorgy Markus *Die Seele und das Leben: Der (Junge) Lukács und das Problem der (Kultur)*

Agnes Heller *Jenseits der Pflicht. Das Paradigmatische der Ethik der deutschen Klassik im oeuvre von Georg Lukács*

Henri Arvon *L'esthétique de Lukács est-elle marxiste?*

Maurice Weyembergh *M. Weber et G. Lukács*

Iring Fetscher *Zum Begriff der (objektiven Möglichkeit) bei Max Weber und Georg Lukács*

VARIÉTÉ

Warren Shibles L'originalité de Wittgenstein

ANALYSES ET COMPTES RENDUS

La philosophie contemporaine: Chroniques. Ed. par Raymond Klibansky (M. Weyembergh) —*Les Études bergsoniennes,* Vol. X. (Ph. Devaux)—Jean Brun, *La nudité humaine* (J. Parain-Vial)—W. K. C. Guthrie, *The Sophists: Socrates* (P. Somville)—V. Descombes, *Le platonisme* (P. Somville)—J. Bernhardt *Platon et le matérialisme ancien* (P. Somville)—R. Hoven, *Stoïcisme et stoïciens face au problème de l'au-delà* (P. Somville)—R. Barthes, *Sade, Fourier, Loyola.* (P. Somville).

OUVRAGES REÇUS VIE PHILOSOPHIQUE
REVUE DES REVUES TABLE DES MATIÈRES DU TOME XXVII (1973)

Abonnements: Belgique: 450 F.B.
Pays européens: 475 F.B.
Autres pays: 500 F.B.

Derniers numéros parus: No. 101: Feuerbach—No. 102: Bertrand Russell
Numéros prévus: Meinong-Lukacs-Methode et Philosophie de l'Histoire

RATIO

edited by Stephan Körner

Volume XVI December 1974 No. 2

RUDOLF HALLER	Concerning the So-called 'Münchenhausen Trilemma'
MORTON WHITE	On the Use of English 'ble'-words to Express Possibility and Necessity
GORDON BAKER	Criteria: A New Foundation for Semantics
M. W. BEAL	Essentialism and Closed Concepts
HILLEL STEINER	The Concept of Justice
JOHN HARRIS	Leibniz and Locke on Innate Ideas
REVIEW:	
MICHAEL WELBOURNE	*The Nature of Things* (Anthony Quinton)

Price £2.25 net **Annual Subscription: £4.00 net**

Orders to: Journals Department, 108 Cowley Road, Oxford OX4 1JF, U.K.

Published by

BASIL BLACKWELL • OXFORD

SYNTHESE

An International Journal for Epistemology, Methodology and Philosophy of Science

Editor: Jaakko Hintikka, Academy of Finland and Stanford University.
Review Editor: Dagfinn Føllesdal, Stanford University.
Associate Editor: Ilkka Niiniluoto, Institute of Philosophy, Helsinki.
Editorial Assistant: Juha Manninen, Institute of Philosophy, Helsinki.

Contents of Volume 26, Nos. 3/4—April 1974

Editorial Note.—Marian Przełęcki / Empirical Meaningfulness of Quantitative Statements. —Raimo Tuomela / Empiricist vs. Realist Semantics and Model Theory.—Rolf A. Eberle/ A Logic of Believing, Knowing, and Inferring.—D. S. Mannison / Lemmon on Knowing.—Morton White / Why Does a Causal Conditional Seem to Assert Possibility When in Fact it Does Not?—John A. Barker / Brand and Swain on Causation.—Marian Przełęcki / On Model Theoretic Approach to Empirical Interpretation of Scientific Theories.—C. A. Hooker / Systematic Realism.—Joseph Agassi / The Logic of Scientific Inquiry.—Announcement: Fifth International Congress of Logic, Methodology and Philosophy of Science.

Subscription price per volume of four issues Dfl. 100,—, plus Dfl. 12,50 for postage and handling. About six issues are published yearly. Personal subscription price on request.

D. REIDEL PUBLISHING COMPANY
DORDRECHT–HOLLAND / BOSTON–U.S.A.

The Monist

An International Quarterly Journal of General Philosophical Inquiry

Founded 1888 by EDWARD C. HEGELER *Editor*, EUGENE FREEMAN

Editorial Board: William P. Alston, Monroe C. Beardsley, Lewis White Beck, William A. Earle, Dagfinn Føllesdal, William Frankena, Maurice Mandelbaum, R. Barcan Marcus, Richard Martin, Mary Mothersill, Joseph Owens, Richard Rorty, J. B. Schneewind, Wilfrid Sellars, John E. Smith, Richard Wasserstrom.

Managing Editor, ANN FREEMAN

EACH ISSUE is limited to articles on a single general topic.
Communicate with the Editor in advance for Special Instructions. Papers should be 6000-8000 words in length and must be submitted in duplicate nine months prior to the scheduled publication of the issue, accompanied by return postage.

GENERAL TOPICS for recent and forthcoming issues:

SCHEDULED PUBLICATION DATES:

Vol. 59,	No. 2	Apr. 1975	Philosophical Problems of Death
Vol. 59,	No. 3	July 1975	Language, Thought, and Reality
Vol. 59,	No. 4	Oct. 1975	The Philosophy of Mysticism
Vol. 60,	No. 1	Jan. 1976	Bioethics and Social Responsibility
Vol. 60,	No. 2	Apr. 1976	Philosophy and Religion in the Nineteenth Century
Vol. 60,	No. 3	July 1976	New Directions in Semantics
Vol. 60,	No. 4	Oct. 1976	Historicism and Epistemology

Editorial Office: Box 1908, Los Gatos, California 95030. Business Office: Box 599, La Salle, Illinois 61301

Subscription Rates: United States: Annual (4 issues): Institutions, $10.00; individuals, $7.50. Single copies: $2.75. Foreign Postage: Add 25 cents to single copy rate or $1.00 to subscription rate.

Philosophy of Science

OFFICIAL JOURNAL OF THE PHILOSOPHY OF SCIENCE ASSOCIATION

Editorial Board:

Richard S. Rudner (Editor-in-Chief), Peter Achinstein, Gustav Bergmann, May Brodbeck, Arthur Burks, C. West Churchman, Theodosius Dobzhansky, Adolf Grünbaum, Edward Madden, Henry Margenau, Gerald J. Massey, Alberto Pasquinelli, Wesley C. Salmon, Kenneth F. Schaffner, Dudley Shapere, J. J. C. Smart, Melford E. Spiro, Patrick C. Suppes, Lewis K. Zerby, Peter D. Asquith (Managing Editor).
Book Reviews: James Van Evra.
Assistant Editors: Patricia Ann Fleming and Paul C. L. Tang.

Vol. 41 March 1974 No. 1

CONTENTS*

The logico-linguistic mind-brain problem and a proposed step toward its solution, Herbert G. Bohnert / *An attempt to add a little direction to "the problem of the direction of time"*, John Earman / *A pragmatic analysis of idealization in physics*, William F. Barr / *Toward a theory of event identity*, A. J. Stenner / *Discussion: Spielman and Lewis on inductive immodesty*, David Lewis / *Discussion: Models, theories and Kant*, A. V. Bushkovitch / *Book Reviews* / *Membership List*.

* The contents of *Philosophy of Science* are indexed in *Science Citation Index*, *ABC Pol Sci*, *Language and Language Behavior Abstracts*, and *The Philosopher's Index*.

Philosophy of Science is published quarterly by the Philosophy of Science Association, 18 Morrill Hall, Department of Philosophy, Michigan State University, East Lansing, Michigan 48823. Membership dues are $10.00 a year and include subscription to the journal. Subscription price for non-members is $15.00 U.S.A. ($16.00 non-U.S.A.). Single numbers $5.00. Subscription inquiries should be directed to the Managing Editor.

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1974-5

President: Professor J. W. N. Watkins
President-Elect: Professor D. J. O'Connor
Hon. Secretary: Dr P. M. Williams
Hon. Treasurer: Dr D. A. Gillies
Members: Dr I. Hacking, Mr H. Martins, Dr M. Machover, Mr N. Maxwell, Mr W. Newton-Smith, Dr E. Zahar

EDITORIAL BOARD OF THE JOURNAL

Professor R. B. Braithwaite
Professor H. Dingle
Professor Sir John Eccles
Professor Sir Harold Jeffreys
Professor W. C. Kneale
Professor S. Körner
Professor M. Polanyi
Professor Sir Karl Popper
Professor G. J. Whitrow
Professor J. H. Woodger

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules :

' The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all persons who have been proposed by two members, approved by the Committee, and elected at an ordinary meeting. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £1·00 ($2.50). Those who wish to join should write to the Hon. Secretary, Dr P. M. Williams, School of Mathematical and Physical Sciences, The University of Sussex, Falmer, Brighton, BN1 9QH.

Brit. J. Phil. Sci. Vol. 25 No. 2 June 1974

The British Journal for the Philosophy of Science

Contents

Articles

99–135 Peter Urbach Progress and Degeneration in the 'I.Q. Debate' (I)

137–154 Roland Puccetti Pattern Recognition in Computers and the Human Brain.

Discussions

155–160 Pavel Tichý On Popper's Definitions of Verisimilitude.

160–166 John H. Harris Popper's Definitions of Verisimilitude.

166–177 David Miller Popper's Qualitative Theory of Verisimilitude.

178–188 David Miller On the Comparison of False Theories by Their Bases.

Reviews

189–206 Jerzy Giedymin on K. Ajdukiewicz: *Problems and Theories of Philosophy*; I. C. Jarvie on W. W. Bartley: *Wittgenstein*; Andrea Mannu on L. Nelson: *Progress and Regress in Philosophy*.

Programme of Meetings

Announcements

Printed in Great Britain at the University Press, Aberdeen

29 JAN 1975

UNIVERSITY * CALCUTTA
CENTRAL LIBRARY

Volume 25 Number 3 September 1974
BJPIA5 25 (No. 3) 209-300 (1974) 0007-0882

The British Journal for the Philosophy of Science

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Acting Editor: JOHN WORRALL

Assistant Editors: COLIN HOWSON AND ELIE ZAHAR

Subscriptions
The *Journal* is published quarterly. The subscription price is £7.00 net (U.S.A. and Canada U.S. $22.00) per volume post free; single issues £2.50 net (U.S.A. and Canada U.S. $8.00) plus postage. The subscription is £4.50 (U.S.A. and Canada U.S. $14.00) for members of the British Society for the Philosophy of Science, who should order direct from the publisher. (Membership of the Society is international and is by no means confined to people resident in the United Kingdom.) Orders may be sent to any bookseller or direct to Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT. Second class postage paid at New York, N.Y. 10001.

Notes to Contributors
Articles, correspondence and books for review should be sent to the Acting Editor, John Worrall at the Department of Philosophy, London School of Economics, Houghton Street, Aldwych, London WC2A 2AE. Articles should be accompanied by abstracts of about 100 words.

Material is accepted for publication in the *Journal* on the condition that copyright is assigned to the British Society for the Philosophy of Science.

Three copies of all contributions should be sent, typed on one side only with wide margins and double-spaced. Footnotes, which should be as few as possible, should be typed separately, double-spaced at the end of the text. They should be referred to by consecutive numbering throughout the text, in which they will be inserted in page proof. Bibliographical references should be listed alphabetically at the end. The style of references should follow current *Journal* practice.

Long quotations are printed in small type without quotes, and should be so marked in the typescript. Single quotes should be used to form the names of words and sentences, and double quotes for all other purposes. Diagrams, tables, and illustrations should be on separate sheets, with their desired position in the text clearly indicated. Symbolic formulae appearing as separate lines should be clearly distinguished in the text, and foreign letters (Greek, German, etc.) indicated in the margin. Material to be printed in italics, especially English letters used as symbols ('*A*', '*x*', etc.), should be underlined. On matters of style and layout, authors should consult CHAUNDY, T. W. *et al.* [1964]: *The Printing of Mathematics*, London: Oxford University Press.

Offprints of articles may be ordered when corrected galley proofs are returned. Fifty are supplied free ; the price of extra copies is given on the order form.

© British Society for the Philosophy of Science 1974

For permission to reproduce material from *The British Journal for the Philosophy of Science*, please apply to the Editor.

Back Volumes
Volumes 1-16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17-24 may be obtained from Cambridge University Press.

Brit. J. Phil. Sci. 25 (1974), 209–221 *Printed in Great Britain*

The Unpredictability of Free Choices*

by G. SCHLESINGER

Received 26th July, 1973

1 Suppose two boxes are placed in front of me and I am given two choices

(1) Take box II only.
(2) Take both boxes.

It is given that there are $1,000 in box I and that a Predictor has either placed $1,000,000 in box II or has put nothing there according to the following rules

(*A*) If Predictor predicts that I will make choice (1) he puts $M into box II;
(*B*) If Predictor predicts that I will make choice (2) he puts nothing into box II.

Now the Predictor is a most remarkable person; on thousands of occasions before he has under all sorts of circumstances predicted correctly and on no occasion has he predicted wrongly the choices of people placed in the same situation as myself. Assuming that I wish to gain maximum amount of money, should I make choice (*1*) or (*2*)?

There is a compelling argument for making choice (*1*). On plain inductive reasoning it is virtually certain that the Predictor has predicted correctly my choice and that if I take both boxes he is sure to have predicted so and has placed no money in box II in which case I gain $1,000 only. On the other hand, if I take nothing but box II it is sure that the Predictor correctly predicted this and hence placed $M in it and so I end up with $M.

There seems, however, also a good argument why I should take both boxes. The Predictor has already made his prediction (say 24 hours ago), he has already either placed $M in box II (and it is guaranteed that once he has put $M there he cannot take it back) or decided not to put anything there. Suppose he has placed $M in box II then by taking both boxes I gain $1,001,000. If on the other hand he has put nothing in box II, then again, by taking both I gain more, namely, $1,000. Thus in either case I increase my gain by taking both boxes.

---

* I wish to thank Stanley Munsat and Ed Galligan who have read and commented upon an earlier version of this paper.

Robert Nozick in his [1969] applies decision theory principles to the problem and after a considerably lengthy discussion concludes that one should take both boxes. Without going into any of the details of his discussion let me state that I believe his whole approach to be misguided. On the assumption that the Predictor is to be believed virtually infallible there is a very simple argument to show that I should make the choice of taking box II only. To let my point emerge more sharply I shall call the previous story Case 1 and shall introduce Case 2:

I have been given the same two choices as before. But now there are $1,000 in box I and nothing in box II. There is an Observer who watches to see whether I take box II only or both boxes and

(*C*) If Observer sees that I have made choice (1) he gives me $M.

(*D*) If Observer sees that I have made choice (2) he gives me nothing.

It is at once clear here that I should take box II only so as to receive $M rather than just $1,000. But why? It seems after all one can develop an argument that I should take both boxes. It is now already certain on the basis of the law of excluded middle, that the Observer will either give $M or he will not. If it is true that he will not give me $M then I should take both boxes so that at least I have $1,000 and if it is true that he will give me $M then too I should take both boxes since it is better to have $1,001,000 than a mere $M.

The second argument, however, is obviously faulty. If I take both boxes I ensure it that I shall get $1,000 rather than $M, which I could gain by making a different choice. True enough $1,001,000 is even more than $M but there is just no possibility for me, no matter what I do, to gain $1,000 as well as $M; for if it is true that the Observer gives me $M it must be the case that I take box II only in which case I do not get the extra $1,000.

Similarly in Case 1: on the surface it may look as if there are four possibilities, namely, that I get nothing, $1,000, $M, or $1,001,000. On a closer look, however, it becomes clear that two of the four possibilities just do not exist. It can never be the case that I end up with nothing because that could only come about if box II was empty and I did not take box I as well; but the infallibility of the Predictor ensures that if I took box II only it was not empty. Nor is it ever possible for me to gain $1,001,000. I could only obtain this amount if I took both boxes and box II contained $M; but the infallibility of the Predictor ensures that if I take both boxes then box II is empty. So there are two and only two possibilities: to gain $1,000 or to gain $M. Obviously I should act so as to realise the second possibility. This implies that I must choose to take box II only.

Maya Bar-Hillel and Avishai Margalit in their [1972], arrive indeed at the conclusion that one should choose to take box II only. Their argument seems to me essentially correct. The trouble, however, is that they arrive at the conclusion that I should take choice (1) and leave matters at that. Matters, however, cannot be left at that since, as I shall show at once, our story involves a contradiction.

Suppose that Smith who is perfectly intelligent and is a perfect well-wisher of mine who always advises me to do what is in my best interest, sits on the opposite side of box II which is transparent but which is hidden from my view. If Smith sees that there is nothing in box II then he surely advises me in his heart that I should take both boxes so that at least I make $1,000. If on the other hand Smith sees that there is $M in box II then he will want me to gain $1,001,000 and thus wishes again that I take both boxes. (If Smith knows about the Predictor then in this case he advises me in his heart to take both boxes but sadly concludes that it is certain that I will not follow his advice.) Smith is not allowed to communicate with me. It does not really matter; I am absolutely certain what he would say to me. How come I do not follow his advice?

Consider the statement: 'Under certain circumstances it may happen that it is not in my best interest to follow the advice of a sufficiently intelligent and well informed perfect well wisher.' If by sufficiently intelligent, perfect well wisher we mean a person who invariably advises me to do what is in my best interest then the statement amounts to saying that sometimes it is not in my best interest to do what is in my best interest to do and is therefore obviously self-contradictory. But exactly the same contradiction is involved in our story. By hypothesis Smith is a sufficiently intelligent and well informed well-wisher. We know for sure that he advises me to take both boxes. But by the argument developed before it has been shown that it is in my best interest not to follow this advice and take box II only. This is a contradiction. It does not really matter that Smith does not exist. The kind of advice he would have given had he existed is known. It is still a contradiction to maintain that it is in my best interest to act contrary to the advice that would have been given by Smith would he have existed.

We are forced therefore to reject our story as an impossible one. But what is impossible? Surely there is nothing contradictory in saying that there exists a Predictor who on thousands upon thousands of previous occasions correctly predicted the choices made by people and that now I am faced with two choices, etc., etc. Now indeed none of the actual facts described in our story are impossible. But our story also contains an assertion which constitutes not a description of an observed fact but which is a conclusion based on inference. The conclusion I am referring to is that the Predictor is virtually infallible as evidenced by his past performance.

This conclusion is, however, not forced upon us by the rules of deductive logic. It is not contradictory to maintain that the Predictor's past successes were not a result of any special ability but were due to an incredibly improbable coincidence and consequently there is no basis to believe that he will have been successful in the prediction he has made regarding my choices.

To maintain such a hypothesis is of course contrary to the tenets of inductive logic. The principles of inductive logic require that the regular association of events in the past is not to be put down to mere chance but be attributed to a lawful connection between those events. The fact that the Predictor has always been successful in predicting people's choices is therefore to be explained as a result of a connection which exists between those choices and his predictions. But if we assign the ability to the Predictor to predict and hence expect that very probably in our case too he has made the right prediction, we are made to act contrary to the advice of a perfectly intelligent and well informed well wisher, which is to be involved in a contradiction. Thus we have the choice either to maintain that the extremely improbable has happened in the past or to maintain something that implies a contradiction. We are forced to choose between going against the principles of inductive logic and going against the principles of deductive logic. Naturally we have to do the former.

Thus in spite of his past record I have to maintain that it is not very probable that the Predictor has correctly predicted my choice. Consequently it cannot be said that by taking both boxes I am making sure that there is nothing in box II or that by taking box II only I can be certain that it contains $M. My choice therefore must be to take both boxes. I conclude therefore that Nozick was right in what choice one ought to make but not for the reasons he gave since he arrived at the conclusion that one should take both boxes in spite of the fact that he entertained no doubts in the powers of the Predictor. I maintain, however, that one should take both boxes because there is no argument left for taking box II only since one is not to believe in the ability of the Predictor to predict.

For a moment it may seem that one could resist our conclusion. Admittedly the statement 'Under certain circumstances it may happen that it is not in my best interest to follow the advice of a sufficiently intelligent and well informed perfect well wisher' is necessarily false. It must therefore be in my best interest to follow the advice of Smith and take both boxes. But in order to explain how this is possible it is not necessary to conclude that the Predictor has no ability to ensure that I get at most a million dollars and at least a thousand dollars, it is possible to deny the credibility of some other proposition based on inductive reasoning. For example it seems possible to maintain that the Predictor is fully capable to ensure that if I take box II only I find a million dollars in it and if I take both boxes I find box II empty, however, Smith is not entirely well informed. Suppose he sees a million dollars in box II. His advice is then that I should take both boxes so that I acquire more than a million dollars. If, however, I were to follow his advice the million dollars would vanish from box II and contrary to what he thought I would end up with a thousand dollars only. True enough in the past this has never happened; whenever the Predictor failed to put money in box II that box remained empty and whenever he did put money there it remained in the box until the latter was opened. It is also given that the boxes are guarded well and it is a known fact that money does not disappear just like that from well guarded boxes. All this, however, amounts to no more than a strong inductive argument why Smith is well informed and why if he sees a million dollars in box II then there is very firm rational basis for believing that by taking both boxes one ends up with $1,001,000. However, we have already agreed that in order to avoid a contradiction we are forced to reject a conclusion based on inductive reasoning. Why then must we reject the conclusion that the Predictor is endowed with special powers why not reject the conclusion that a million dollars placed in box II does not spontaneously disappear?

The answer is that indeed all we know for sure from the fact that we have to avoid a contradiction is that some inductively arrived at conclusion must be given up but we are not directly given which one of such conclusions must be rejected. In such a situation it is obvious that we are to give up the conclusion which is most reasonable to give up, that is the conclusion which is not as strongly supported as any other and also for which we can produce the best explanation why it should be rejected. The conclusion that the Predictor is endowed with a special ability is based on the evidence of his past performance which of course may be very extensive. On the other hand the conclusion that once a million dollars are put in box II they do not vanish is based not only on our past experience with all the instances of this particular game but on a vastly wide range of cases in the past in all of which items which were enclosed in well guarded containers did not spontaneously disappear. In addition to this, one would be hard put to produce a reasonable explanation why such a phenomenon as the disappearance of money from a closed box should occur. On the other hand as we shall soon see there is a reasonable explanation why the Predictor should be powerless to foresee the choice I am about to make.

Another attempt to escape the contradiction while at the same time maintaining that the Predictor does have the power to ensure that his plans are carried out would consist in claiming that in case there are a million dollars in box II, I am powerless to take both boxes.

Of course such a possibility exists, but it should not deflect me from the course of action I must choose. I must attempt to follow the advice of Smith and take both boxes. Should I find this physically impossible then naturally, against my will, I shall be forced to take box II only. It will, however, become clear to me that the only way for the Predictor to ensure the success of his prediction is by forcing me to comply with it. The essence of the conclusion we have arrived at earlier remains true just the same: the Predictor has no ability to predict what my freely willed choice is going to be.

Finally it should be mentioned that sometimes Newcomb's story is presented not merely with a Predictor who has a vast record of successes in predictions under all sorts of conditions but simply with a Predictor who is postulated to have the ability to predict with absolute certainty what an agent's choice will be. But what we have shown is that we just cannot postulate that anybody has such ability, because of the contradiction it leads to. Of course it cannot be denied that it is conceivable—though incredibly improbable—that someone should have an actual record of successes so extensive that the rules of inductive logic should require us to assign an ability to him no less than those rules require us to subscribe to other well established empirical generalisations. But in this very unlikely situation we would be required to refuse to follow the tenets of inductive reasoning in order to avoid a contradiction.

2 Our story has thus taken a surprising turn; we have been presented with the case of a man who seems to have demonstrated thousands upon thousands of times that he possesses infallible powers of predicting the choices that are going to be made, yet we are to ignore the evidence of our eyes and contrary to the universally binding principles of inductive logic declare his past successes as due to nothing more than exceedingly improbable accident.

We have then here an instance of a simple human act, that of choosing between taking both or just one of the boxes available and it has been demonstrated that this act is essentially unpredictable not only on the basis of the knowledge of all the laws governing human behaviour and the full knowledge of the present state of the agent, but unpredictable also by direct precognition or by any other method. Are we not entitled to draw from this far reaching conclusions concerning the fundamental freedom of the will?

Before answering this question we should be reminded that without using Newcomb's problem it can be shown, and has been shown, that there are cases in which it is impossible to predict with certainty what a voluntary act will be. The point of this section will be to argue that all those other cases in which prediction fails are special cases from which it may be impossible to infer that a normal voluntary act is unpredictable in principle. In our case, however, there is nothing extraordinary in the circumstances of the agent and it should therefore provide a basis for inferring from it that in general human choices are unpredictable.

One well known case is where before acting the agent is informed how he was predicted to act. It is clear that in this case there can be no guarantee that the prediction will come true since the agent knowing what he was predicted to do has the option to do the opposite. But from this special case it does not seem warranted to draw conclusions about the general case. After all there is a fundamental difference between communicated and uncommunicated prediction. One might perhaps be tempted to argue this by claiming that the failure of communicated prediction is simply due to the fact that the prediction has not taken into account all the factors influencing the agent. When a prediction is made how an agent will act and the agent is told about this then the prediction may have been made by taking into account all the agent's circumstances but the extra fact that the prediction is disclosed to him was not taken into account. But that extra fact changes the agent's state and may induce him to act differently. Of course the predictor may go a step further and predict$_2$ what effect the disclosure of prediction$_1$, based on the agent's nature and circumstances, will have on him. As long as he does not disclose to the agent this latter prediction$_2$, it may well be that his prediction$_2$ is guaranteed to come true. If, however, he communicates to the agent prediction$_2$ as well then he may fail in that prediction too since now again the agent's circumstances are different from those upon which prediction$_2$ was based: prediction$_2$ has been made on the basis of all the initial conditions of the agent including the fact that prediction$_1$ had been revealed to him. The agent's action, however, is now also influenced by an additional factor, namely his knowledge of prediction$_2$. One can go on like this and it will be general be true that prediction$_n$ can take into account all the circumstances of the agent including his knowledge of the predictions which were made about him up to prediction$_{n-1}$. However it will not have taken into account the effect of the disclosure of prediction$_n$ itself. Nothing therefore can be inferred from the failure of communicated prediction to the failure of prediction in general since the former is based on essentially deficient information.

Whatever the worth of the foregoing argument, it surely does not apply to predictions which do not require the taking into account the state of the agent prior to his choice, but parapsychological predictions which may be based on directly perceiving now how the agent is going to choose to act. Essentially therefore the reason why from the failure of communicated prediction we may not infer the failure of ordinary prediction is because the two are made under basically dissimilar conditions. The circumstances of the agent in the former case include a very significant factor absent in normal cases of choice. The decision of the agent in cases of communicated prediction is crucially influenced by his knowledge of how he was predicted to act. In that case therefore there is a compelling reason why the prediction cannot be guaranteed to come true. It is unwarranted to conclude that in normal cases where such a reason does not exist the agent's actions are also unpredictable in principle.

The situation is largely similar in the case of a very intelligent agent who for some reason is determined that his actions should not be predicted and a predictor who attempts to make his predictions on the basis of his knowledge of the habits and inclinations of the agent. The agent being acquainted with his own nature may well succeed in figuring out what the predictor reckons he is going to do and do the opposite. In the contest of wits between the predictor and the agent the latter has a good chance to defeat the former no matter how far the science of psychology advances and how perfected our knowledge of human nature might become. As the science of psychology advances the predictor will have stronger and stronger basis upon which to try to foretell the actions of the agent, at the same time, however, the agent too, will be able to avail himself of everything that is at the disposal of the predictor and figure out his prediction and frustrate him.

The reason why the agent will often prevail over the predictor is because in the situation described we have the essential features of a communicated prediction. But as we have already said we must not draw general conclusions from communicated predictions, where compelling reasons exist for the failure of prediction, to cases where such reasons do not exist.

Another instance in which prediction seems impossible is provided by our Case 2 discussed in the previous section. Here it still seems clear that the reasonable choice for me is to take box II only so as to ensure that the Observer gives me $M. But for a moment we may hesitate. Suppose we have Smith's counterpart, Jones, who is also a well wisher of mine with strong precognitive powers who, looking into the future can now observe whether or not I am eventually to receive $M from the Observer. Suppose Jones sees that I will not receive $M. Then, being a well wisher, he surely wishes that I take both boxes so that I gain at least $1,000, and if he sees that the Observer gives me $M then again he advises me in his heart to take both boxes so that I gain an extra $1,000. Jones is an intelligent, well informed well wisher—how can I act contrary to his advice of which I am fully aware without his actually disclosing it to me?

The answer must be that I am not to believe that Jones is well informed. Just as in Case 1 we denied that the Predictor had actual powers of predictions, here too we deny that Jones has such powers. We shall allow no amount of prior success in the past to convince us that Jones has the ability to foresee what the Observer will actually do. His past record, however extensive and however consistent will be put down to extraordinarily improbable good luck.

But Jones's inability to predict correctly what the Observer will do constitutes once more a special case from which we may not draw general conclusions. Suppose Jones 'sees' that the Observer gives me $M and instructs me to choose both boxes. His advice is of course designed to secure me a larger amount of money but in fact it is equivalent to an advice by Jones himself for me to act contrary to his own prediction. For Jones's 'seeing' that I receive $M amounts to his 'seeing' me choosing box II only, since it can be taken as absolutely certain that the Observer will give me $M if and only if I take box II only, for there are no obstacles in the Observer's path to keep strictly to his conditions. But Jones explicitly advises me to act contrary to his prediction, and if I follow his advice—which I cannot afford to ignore if I am to assume that he is well informed—then I take both boxes which is to act contrary to how I was predicted to act. But from the failure of prediction in this special case where the circumstances surrounding my choice included my having been advised to act contrary to how I was predicted to act cannot serve as a basis for inferring from it the failure of prediction under normal circumstances.

It may perhaps seem, however, that one could raise the following question: suppose Jones 'sees' that I do not get $M and advises me to take both boxes and I ignore his advice and take box II only. Here it is not the case that I was acting upon the advice of Jones who was instructing me to act contrary to his own prediction. Here Jones advises me to act in accordance with his prediction, but I ignore the advice and do not take both boxes ending up of course receiving $M since the Observer will act according to what I actually do and not according to what I was told to do by Jones. What special explanation exists in this instance for Jones's failure to envision the future correctly?

But of course there is absolutely no need to assume that if Jones 'sees' that I do not get $M then his vision is not going to be fulfilled. It is possible to maintain that whenever I do take box II only the Jones's will have 'seen' that I do get $M (except that even in this case his advice still remains 'take both boxes').

Thus the circumstances surrounding my situation in Case 2 are such that it cannot serve as a basis from which one could infer the general unpredictability of choices. In those cases in which Jones may fail to predict correctly whether the Observer gives me $M there are special reasons—absent in normal cases—why Jones may fail. On the other hand in those cases in which there is no special reason why Jones might fail he never actually fails. Jones may be unsuccessful in predicting correctly what will happen if I take both boxes in which case he may have predicted that I *do* get $M, but his failure is explainable because he himself has advised me to act contrary to his prediction. In cases where I take box II only, there is no special reason why Jones's prediction should fail to come true but neither can it be established that they do fail.

So far then in all those cases in which we have seen that predictions concerning human choices cannot be relied upon, it has been demonstrated that the choice was made under special circumstances which provided clear reasons—not existing in general—why the prediction should not be reliable. Does the failure of the Predictor in Newcomb's case introduce an essentially novel situation with far-reaching implications or are there special reasons in this case also which prevent prediction? It might seem that once more we have a situation in which my actions are unpredictable for the special reason that I am acting on the advice of Smith who knows, by looking at the contents of box II, what the Predictor has predicted and advises me to act so as to frustrate him. But this is not true if I take box II only, in which case I disregard, for whatever reason I may have, the advice of Smith who is in a superior position to advise me, and thus we can cite no special reason at all why the Predictor should fail to predict that I take box II only.

Perhaps it will indeed be suggested that if I do take box II only, it can be predicted, if not with certainty, at least with high probability that I will do so. It can, however, be shown easily, that whatever the probability that the Predictor predicts that I will take box II only, this probability cannot be higher at all in case I actually do take box II only than if I take both boxes. For suppose to the contrary that the probability of the Predictor predicting that I take box II only, if I end up taking both boxes is p, while if I in fact take box II only then it is $p + \epsilon$. Let the amount of money the Predictor places in box II, if he predicts that I will take that box only, x dollars rather than a million dollars. The expected desirability of choosing both boxes is $1{,}000 + px$, since if I take both boxes I am sure to receive $1,000 and there is a probability of p that in addition I also receive x dollars. The expected desirability of choosing box II only, is, on the other hand $(p + \epsilon)x$. By making x sufficiently large ϵx can be made to exceed 1,000, no matter how small ϵ is. The expected desirability of choosing box II only is higher than that of choosing both boxes and therefore it is in my best interest to choose box II only. But this is contrary to the advice of Smith and is therefore impossible.

We have thus seen that even in those cases in which I choose to take box II only the frequency with which the Predictor will predict this will not increase over the frequency with which he makes the same prediction in cases in which I do not choose to do so. Now in this case, however, there is nothing special in my circumstances which tends to explain why here, but not in general, there is no greater success in predicting my choice. Our argument employing Smith and the consequent showing that there is a contradiction if we assume that the Predictor can predict with higher probability than p (where p is the probability that he will predict that I choose box II in case I do choose to take both boxes) that I will choose box II only if I do in fact choose box II only is certainly not an argument *why* the prediction fails to be more successful but merely an argument *that* it fails to be so. If we want to know why the prediction is not more secure we must look for the reason in the general nature of free human choices. It may perhaps be said that it is of the essence of a free agent's voluntary choices that they are not determined at all before the very moment he wills his decision and that is why there is no way to foresee it.

It should be pointed out that it is not claimed that it is an even chance that an agent in Newcomb's case will make either choice. It may well be that it is highly probable—especially if our solution according to which it is rational to choose both boxes—that the agent will choose to take both boxes. Consequently if the Predictor wants to increase his chances to be right he should predict that the agent will take both boxes. What is being claimed is, that the Predictor will have no success at all in sorting out the minority cases in which the agent chooses box II only. The frequency with which the Predictor predicts that the agent will take both boxes will not vary from cases in which the agent actually takes both boxes to cases in which he takes box II only.

It should perhaps also be pointed out that it is not claimed that the Predictor may not be able to distinguish between agents with different dispositions. Suppose I am an abnormal person with a strong psychological compulsion, whenever presented with choices like in Newcomb's case, to take box II only. It may therefore be highly probable in my case—but not in the case of normal persons—that I will choose to take box II only. There is no reason why the Predictor should not be aware of my affliction and consequently make a different forecast in my case than he would make in other cases. The important point, however, is this: I may not always take box II only, but only in the majority of cases and the Predictor will have no ability at all to sort out the minority cases. This means that in those instances in which I do choose both boxes he is not in the least more likely to predict that I do so than when I do choose box II only. In other words, since my choices are not entirely unconstrained because of the presence of the psychological compulsion the Predictor may utilise this and predict most of the time that I will take box II only and be most of the time right. But in the minority of cases in which I exercise my will to overcome my compulsion, the frequency with which the Predictor will still predict that I take both boxes will not decrease at all.

3 In Section 1 I compared Case 1 to Case 2 with respect to the choice I should make. Now we have seen, however, that the two cases are essentially different. In the former it is reasonable for me to take both boxes; in the latter, box II only. The difference arises out of the fact that in Case 1 it is a Predictor offering $M on certain conditions whereas in Case 2 it is an Observer. There are no obstacles for the Observer to adhere strictly to his condition to give me $M if I take box II only and to give me nothing if I take both boxes. Since having $M is preferable to having $1,000 only, obviously I should take box II only. The Predictor on the other hand is powerless to ensure that everything happens according to his conditions. He may predict that I take both boxes and hence not put anything in box II yet I take eventually box II only and end up with nothing; it is also possible that he predicts that I shall take box II only and thus put $M in it yet in the end I take both boxes, thus gaining $1,001,000. Since we do not believe in his power to predict there is no reason to think that by taking both boxes it becomes more likely that there is money in box II or that by taking box II only, I render it more probable that there are $M in it. Obviously, therefore I should take both boxes which contain more money than box II alone.

There are other differences as well between the two cases. In both cases we have perfect well wishers whose advice to me is to take both boxes. In Case 1, to be rational, I must accept the advice of Smith who is a sufficiently well informed well wisher and whose advice is therefore in my best interest to follow. In case 2, however, I reject Jones's advice on the grounds that he is not well informed. Smith must be well informed since the only relevant information he requires is whether there are or there are no million dollars in box II and there are no obstacles preventing him from observing this. Jones on the other hand bases his advice on what he 'sees' the Observer will do. His information may not be reliable, for if he 'sees' that the Observer gives me $M and I follow his advice taking both boxes, it will of necessity turn out that contrary to Jones's 'vision', the Observer, who is in a solid position to determine that he follows his conditions, will not give me anything.

In Case 1, however, even in those instances in which I disregard the advice of Smith and take box II only the Predictor may fail to foresee that I will do so. Furthermore the frequency with which the Predictor will predict that I take box II only is not in the slightest greater in those instances in which I actually do so than the frequency with which he makes the same prediction when I do otherwise. But there is no special reason why the Predictor should have no greater success in these instances. It seems therefore that the frequency of the failure of the Predictor in these instances points to a fundamental aspect of voluntary choices namely their unpredictability. In general when I can choose between doing A or refraining from doing A then while it may be possible to predict that more probably I will do A than not, in case where there is a small probability that I may refrain from doing A it is not possible to be successful in guessing this. That is, it is not possible that a predictor will predict that I do not do A with greater frequency in cases in which I actually refrain from doing A than in cases in which I do A. Because of their fundamental unpredictability, voluntary choices may therefore perhaps be said to be free in a very radical sense.

The University of North Carolina
at Chapel Hill

REFERENCES

Bar-Hillel, M. and Margalit, A. [1972]: 'Newcomb's Paradox Revisited', *British Journal for the Philosophy of Science*, **23**, pp. 295–304.

Nozick, R. [1969]: 'Newcomb's Problem and Two Principles of Choice' in N. Rescher (*ed.*): *Essays in Honor of C. G. Hempel*.

Brit. J. Phil. Sci. 25 (1974), 223–234 *Printed in Great Britain*

Propositions, Possible Languages and the Liar's Revenge

by JOHN F. POST

An adequate theory of natural language should include a solution to the semantic paradoxes. This is not the only condition of adequacy, but it is an important one. For the paradoxes are among the crucial tests of one's theory of natural semantics, syntax and phonetics, all of which figure in the paradoxes to some degree. The theory should either show how a correct conception of our language would block the derivations of paradox, or explain why they cannot be blocked even on a correct conception, and justify a relatively simple linguistic reform which would do so.

Just as a theory of natural language should afford a solution to the paradoxes, one's solution typically will presuppose a conception of language. Hence much more is at stake than sometimes meets the eye when philosophers argue about classical paradoxes such as the Liar, the Grelling, and so on. It is simply not true that these arguments are 'five-fingered exercises' or 'sterile ornaments' of a philosophy of language (as I have heard them dismissed by two prominent philosophers, who in fairness should remain unnamed, not having committed themselves on the point in print). One's treatment of semantic paradox can crucially affect one's views about meaning, truth, propositions, sentences, and utterances, together with their interrelations and much else.

This fact may occasionally be obscured by the belief that the classical paradoxes have long since been solved or quarantined by some means or other. But even if they have been adequately solved—which is far from clear—there is no guarantee that the same means would suffice for any new paradox that might turn up. Of course a sweeping rejection of all self-reference has nothing to fear from a new paradox. But natural-language solutions typically make a virtue of allowing some self-reference, in contrast, say, to a strict language-level approach. Furthermore, all such solutions have concentrated on absolute notions of truth and falsity, as expressed in one-place predicates like 'is true' and 'is false'. Relativised predicates like 'is true (false) *in possible world W*' and 'is true (false) *with respect to language L*' have been neglected. Yet paradox lurks in this neglected portion of the expressive power of natural language.

Received 1 *December* 1972

One such paradox is the Liar's Revenge. This arises in connection with sentences like 'What I am saying is not true *in the nether world*', where the nether world is a logically possible world we may imagine the Liar to have joined when he departed this one. Elsewhere [1973] I argue that what is said in 'What I am saying is not true in the nether world' can be shown to be both true and false, given assumptions having at least the same initial plausibility as those used to derive inconsistency in the case of the Liar. There is the considerable danger that some leading solutions to the paradoxes could not solve the Liar's Revenge even if they happened to be entirely successful in solving the Liar and other paradoxes hitherto available. This is indeed the fate of category solutions and state solutions, or so I argue in my [1973]. Such solutions try to block derivation of paradox by pointing to a category mistake in some step of the derivation, or to an incorrect use of a sentence about states of affairs.

In what follows, I argue that a linguistic analogue of the Liar's Revenge poses an even greater threat to proposition solutions. Such solutions posit propositions as truth-bearing constants of translation, and require that any reference to a proposition must in principle be replaceable by its expression. The threat to such solutions is not merely that they cannot solve the linguistic analogue, and therefore are incomplete. What is much worse, basic features of these solutions actually give rise to the linguistic analogue, so that they are inconsistent. Such a *reductio* would be of considerable interest even if, as I happen to believe, the literature already contains logically weaker forms of criticism which proposition solutions cannot answer satisfactorily.

There is another reason, and a deeper one, for continuing to probe proposition solutions (henceforth, '*PS*'s' for short). The Liar's Revenge bears on the question of what counts as a possible world. (*Cf.* Martin [1974] and Post [1974].) Its linguistic analogue bears on the question of what counts as a possible language, as we shall see. (*Cf.* Quine [1960], pp. 213–14; Herzberger [1970], pp. 147, 161; and Chomsky [1965], chapter 1.) *PS*'s presuppose an answer to this question which actually gives rise to paradox. They presuppose a familiar conception of what sort of thing a language is and what sorts of things can be expressed in a language. Given this conception, not only is a certain language possible—we shall come to call it Netherese—but we may plausibly say certain things about Netherese in English. In particular, we may say such things as 'With respect to Netherese, my present utterance does not express (translate into) a true proposition', or briefly, 'What I am uttering is not true in Netherese'. English sentences such as these represent the linguistic analogue of the Liar's Revenge. Basic features of *PS*'s imply that their utterance would express a proposition. The proposition turns out to be both true and false. Evidently there is something incoherent in the underlying conception of what counts as a possible language and what we can say about it in the possible language we call English.

In the concluding section, I offer a tentative diagnosis of the source of the trouble. It does not turn simply on the positing of propositions. As we shall see, the paradox goes through for sentences as well. Rather it turns on the positing of propositions in such a way that certain languages are possible and certain translations are warranted from one possible language to another. If sentences (or sentence-types) are posited in such a way as to allow the same translations and possible languages, then in effect they behave as propositions. The term '*proposition* solution' would still apply, even if the talk is overtly of sentences, so I use the term throughout.[1]

1 There is a familiar conception of natural language according to which there are propositions and there are sentences whose utterance may or may not express a proposition. The utterance of different sentences in different languages can express one and the same proposition. This is a standard way of construing translation: propositions are constants expressed in common by foreign utterances and their domestic translations. In the same vein, because there is no necessary but only accidental connection between sound and meaning,

(*a*) One and the same utterance could happen to express distinct propositions in different languages.

Furthermore, if the *proposition* expressed by an utterance U according to the conventions of a language L is true, then we say *derivatively* that U is true in or with respect to L. (*Cf.* Kneale [1972], p. 239.) Since the same utterance could express one proposition in L and another in L^*, and since both propositions, though distinct, could happen to be true, it follows that

(*b*) Utterances which are true with respect to one language could also be true, coincidentally, with respect to another.

That is, different languages L and L^* could happen to map one and the same utterance U onto distinct propositions in such a way that if U is true in L then U is true in L^*. Jokes about fractured French and fractured Latin depend upon essentially this possibility. Principle (*b*), like (*a*), reflects a standard view of translation. There are propositions and there are sounds or utterances. The truth-values and relations among the former are absolute, in no way dependent upon their accidental connections with the latter. Translation merely finds these connections, without affecting the truth-values or relations.

---

[1] I am heavily indebted to Mr J. L. Mackie for criticism of an earlier draft; and to Professor R. L. Martin for a remark which set me thinking about possible languages in connection with the Liar's Revenge. Naturally, neither is thereby implicated in the outcome. Mackie ([1973], p. 253) presents an interesting 'variant of the liar expressed in terms of languages', to support an objection to Tarski.

In addition, the conception of language at issue holds both that facts about language such as (*a*) and (*b*) can themselves be expressed in natural language, as I just did, and that natural expressive power includes a high degree of self-reference, though not so much as to encompass paradox. As a leading proponent of *PS*'s says, 'In any fully human language it is possible to talk about talk and about what is said in talk'—in that language itself or in some other. (Kneale [1971], p. 327.) Hence 'any human beings . . . have or can easily make a way of talking about propositions', so long as any such talk about a proposition is replaceable in principle by its expression. (Kneale [1972], pp. 233, 235.)

In particular, a natural language at least can express its own phonetics and syntax, even if there should happen to be problems about expressing all of its own semantics. Among other things, this means that given any utterance *U*, then any 'fully human language' *L* will be such that

(*c*) *L* contains sentences whose utterance U^* would express in *L* the proposition that *U* is (or is not) an utterance of a sentence of *L*, or of some other language.

Thus suppose someone speaking Chinese utters *U*. English contains sentences whose utterance U^* would express in English propositions to the effect that *U* is not of a sentence of English, that *U* is of a sentence of Chinese, and so on. Principle (*c*) holds even when *U* and U^* are identical, as in certain cases of self-reference. For example, English contains the sentence 'My present utterance is (is not) an utterance of an English sentence'. An utterance U^* of this sentence expresses in English the proposition that my utterance (call it *U*) is (is not) an utterance of an English sentence. Here $U^* = U$. (*Cf.* Kneale [1972], p. 242.)

Principles (*a*) – (*c*) are at least implicit in ordinary talk about languages, utterances, propositions and truth. In fact, the ordinary, common-sense nature of this conception of language constitutes much of its justification in the eyes of those philosophers who rely on it to help solve the paradoxes. All the more surprising, then, that the conception itself gives rise to paradox. It does so in part because it implies that a certain language *L* is possible, where to say that *L* is possible, given principles (*a*) – (*c*), is just to say that *L* satisfies (*a*) – (*c*).

2 The Liar and his colleagues in the nether world have discovered that they require a language in common in order to pursue their logical investigations. Over the centuries a Netherese language has evolved, which

satisfies principles (*a*) – (*c*) in the following particular ways. To begin with,

A. One and the same utterance expresses distinct propositions in Netherese and English, if it expresses any at all.

A merely specifies that Netherese and English satisfy (*a*), as do Chinese and English, for example, or even Netherlands Dutch and English.

As for principle (*b*), it turns out that certain utterances which are true with respect to English happen also to be true with respect to Netherese. Specifically, the Liar and company have found that every utterance *U* which expresses in English a true proposition about utterances translates, by sheer coincidence, into a true proposition in Netherese. By *A*, of course, the two propositions are distinct. Examples of propositions about utterances would be that Nelson's last words are famous, that the King's latest utterance was melodramatic, that nevertheless it expressed a true proposition in French, that it was not an utterance of an English sentence, and so on. Thus what the Liar has discovered is that where *U* expresses in English a proposition about utterances, if *U* is true in English then coincidentally *U* is true in Netherese. This conditional will be abbreviated, using Kneale's own notation ([1972], p. 239), as

B. $T_E(U) \supset T_N(U)$.

B merely specifies a particular way in which *N* (*i.e.* Netherese) satisfies principle (*b*). Like (*b*), *B* reflects a standard view of the translational enterprise.

In accordance with principle (*c*), *N* contains grammatically well-formed sentences whose utterance U^* would express in *N* the proposition that a given utterance *U* is (or is not) an utterance of a sentence of *N*, or of some other language; English for instance. The sentences of *N* whose utterance happens to express these matters of phonetics and syntax are the formulas '*U* is true in *N*', '*U* is not true in *N*', and '*U* is true in English'. Obviously these formulas are also well-formed sentences of English. We abbreviate them as '$T_N(U)$', '$\sim T_N(U)$', and '$T_E(U)$', respectively, where '*U*' is a stand-in for any name or description of an utterance. But by the conventions of *N*,

C. Utterances of '$T_N(U)$', '$\sim T_N(U)$' and '$T_E(U)$' express the propositions that *U is of a sentence of N*, that *U is not of a sentence of N*, and that *U is of a sentence of English.*

In line with *A*, the propositions expressed in *N* by uttering these sentences are quite distinct from the propositions expressed in English. For example, the proposition that the utterance *U* is not of a sentence of *N* is quite distinct from the proposition that *U* is not true in *N*. The former proposition has to do with the phonetics and syntax of *N*, the latter with the semantics as well. Yet both are expressed (one in N, the other in English) by uttering the sentence '$T_N(U)$', or the sentence it abbreviates.

A further example seems in order, to illustrate the operation of B and C. Consider an utterance U of 'The President's latest utterance is not true in N'. Suppose U expresses a proposition in English, as normally it would, namely the proposition that the President's latest utterance is not true in N. Clearly this proposition is about utterances, namely the President's latest. Suppose further that this proposition is true, say because the President was not speaking Netherese, so that his latest utterance was not even of a sentence of N, and therefore could not have been true in N. Then the proposition U expresses in English is true, which is to say that $T_E(U)$. In light of B, therefore, $T_N(U)$—*i.e.* the utterance U of 'The President's latest utterance is not true in N' expresses a true proposition in N. The proposition expressed, according to C, is that the President's latest is not of a sentence of N, which is true enough given that the President was not even speaking N. *Had* he been speaking N, his latest *would* have been of a sentence of N, so that the proposition expressed by U in N would be false. Hence it would be false that $T_N(U)$, and by B, therefore, false that $T_E(U)$, and U would not be true in English, either because the proposition U expressed in English is false—the President's latest *is* true in N—or because for some reason U failed to express a proposition in English to begin with.

According to a leading proponent of PS's, 'The rationality of man . . . is surely bound up with his use of language . . . and language makes us rational only in so far as it enables us to conceive possibilities . . .' (Kneale [1972], p. 225). Such possibilities include other languages. Is N a possible language? Given the conception of language represented in principles (*a*) – (*c*) it certainly is. Conditions $A - C$ merely specify particular ways in which N satisfies the principles (*a*) that the same utterance can express distinct propositions in different languages; (*b*) that utterances true in one language may coincidentally be true in another; and (*c*) that any fully human language enables us to talk about talk, and thus contains sentences whose utterance would express, for any given U, the phonetic-syntactic proposition that U is (or is not) of a sentence of that language, or of some other. So (*a*) – (*c*) are jointly satisfied by N. It follows that N counts as a possible language given the conception of language presupposed by PS's. Next we shall see how a contradiction can be derived from this conception.

3 Having analysed Netherese, the Liar is having second thoughts about what he said while among the living. It now occurs to him that he ought to have uttered

(*1*) My present utterance is not true *in Netherese.*

If we let '*k*' designate the utterance I produce when I utter (*1*), then (*1*) may be abbreviated as

(*K*) $\sim T_N(k)$.

In effect (*K*) reads, 'My present utterance *k* does not express a true proposition with respect to *N*', or briefly, 'My utterance is not true in *N*'.

In the next section we shall see why an utterance of (*1*), or of (*K*), does express a proposition in English, in spite of the self-reference. At least it does so according to *PS*'s and the general conception of language lately explored. But for the time being we shall merely assume that my utterance *k* of (*1*) expresses a proposition in English. Note that the proposition expressed by *k*—the proposition that *k* is not true in *N*, or that $\sim T_N(k)$—is about utterances, namely *k*. Hence *B* applies, so that *if* $T_E(k)$, then $T_N(k)$.

The question arises as to whether the proposition that $\sim T_N(k)$ is true. Obviously it is. By the conventions specified in *C*, *k* expresses in *N* the proposition that

(*2*) *k* is not an utterance of a sentence of *N*.

But precisely because *k* expresses a proposition in *N*, *k is* an utterance of a sentence of *N* (*cf. C*). So the proposition expressed by *k* in *N* is not true but false. Hence *k* is not true in *N*, which is to say that $\sim T_N(k)$. However, because $\sim T_N(k)$, it follows by *B* that $\sim T_E(k)$, which is to say that *k* does not express a true proposition in English (*cf.* what happened when the President *was* speaking *N*). But by hypothesis, *k* expresses a proposition in English, namely the proposition that $\sim T_N(k)$. It follows that this proposition is not true but false, which entails that $T_N(k)$. That is, *k is* true in *N*. Thus we have proved a contradiction: $T_N(k)$ and $\sim T_N(k)$.

This is the linguistic analogue of the Liar's Revenge. One cannot conclude from the contradiction simply that no such language as *N* could exist, as one concludes in the case of the Barber simply that he could not exist. As we have seen, the key properties of *N*, unlike those of the Barber, are deeply embedded in ordinary talk about languages, utterances, propositions, and truth. Principles (*a*) – (*c*), like the conception of language they represent, are based on that talk; and *A* – *C* merely specify particular ways in which *N* satisfies (*a*) – (*c*). Rejecting the possibility of *N* would require us to reject or revise something in (*a*) – (*c*)—a counter-intuitive and painful business at best, certainly from the point of view of ordinary talk and of *PS*'s, which rely on that talk. Rejecting the possibility of the Barber costs nothing by comparison. So the contradiction represents a genuine antinomy, not a 'veridical paradox'. (*Cf.* Quine [1966], p. 5.) Even if the contradiction did not represent an antinomy, it would represent a reduction to absurdity of the conception of language which gives rise to it.

4 In deriving the contradiction, we assumed that my utterance *k* of (*1*), or of (*K*), expresses a proposition in English. The assumption not only is not blocked by *PS*'s, it is implied by them. It is not blocked because the principle used by *PS*'s against paradoxes like the Liar has no effect here. The principle is that any ***designation*** of a proposition—any name or description—must be replaceable by its ***expression***. The same proposition must therefore be designatable by prefixing the nominaliser 'that' to one of its specimens. (*Cf.* Kneale [1972], pp. 233, 239, 242.)

This is easily done not only for '$\sim T_N(k)$' but for '$T_N(k)$'. These formulas abbreviate '*k* does (does not) express a true proposition with respect to *N*'. The latter contains only one reference to or designation of a proposition, namely one purportedly expressed by *k* in *N*. Since *k* is an utterance of '$\sim T_N(k)$', the alleged proposition is one expressed in *N* by an utterance of '$\sim T_N(k)$'. According to the conventions of *N* specified in *C*, this is the proposition that (*2*) *k* is not an utterance of a sentence of *N*. (*2*) has to do solely with phonetics and syntax; it contains no reference to a proposition, but only to utterances (including *k*) and sentences of *N*. Hence nothing about *PS*'s prevents us from saying that an utterance of (*2*) expresses a proposition in English, or that the proposition can be designated by prefixing 'that' to one of its specimens, namely to (*2*). Even when construed self-referentially, an utterance of (*2*) expresses a proposition (cf. the remarks following (*c*) in section 1). The principle that 'ability to *express* a proposition can never depend on ability to *designate* it' has not been violated; no state of affairs has been presented only by talk about its presentation. (Kneale [1972], p. 243.)

The point merits repeating in more general terms. According to *PS*'s, as seen, 'in any fully human language it is possible to talk about talk and about what is said in talk'—in that language ***or in some other***. One who utters '$\sim T_N(k)$' in English talks about talk—namely about his utterance *k*—and about what is (or perhaps is not) said (expressed) by that talk ***in Netherese***. Even when the English speaker's utterance contains a reference to itself, he may still talk about that utterance and about what it expresses in *N* or in some other language besides English. What he cannot do in a given utterance *U* is talk about what *U* itself expresses in English. In that case the only specimen of the alleged proposition he could produce would be *U*. But then it would become impossible to express the proposition independently of designating it. The English speaker would fall into the same 'vicious circle of preconditions' that blocks the Liar, any time he tries to talk about what his present utterance expresses in English. (Kneale [1971], pp. 330–1, and [1972], pp. 242–3.)

The situation is very different when the English speaker talks about a proposition his present utterance U coincidentally expresses in a language L *other* than English. In that case he *can* go on to produce a specimen of the proposition which is quite distinct from U. At least he can do so if the conventions of L are appropriately different from those of English, so that the proposition that-P^* which U expresses in L is distinct from the proposition that-P which U expresses in English. For then he can produce *in English* an utterance U^* which is distinct from U, and which expresses that-P^* (just as an utterance of (*2*) is distinct from *k;* so too for the corresponding propositions). Hence he can go on to express the proposition that-P^* independently of designating it, simply by uttering U^* in English (*e.g.* utter (2)).

This arrangement also satisfies related Rylean requirements for namely-riders. (*Cf.* Ryle [1951].) For example, I succeed in stating something when I utter 'My present utterance U expresses a true proposition in N', because I can go on to say, '. . . namely the proposition that U is of a sentence of N'. My statement unpacks in such a way that its namely-rider is fulfilled by something other than my statement.

That my utterance k of '$\sim T_N(k)$' expresses a proposition in English not only is not blocked but is implied by basic features of *PS*'s. Perhaps the most basic feature is that whether an utterance expresses a proposition depends on a variety of contextual matters, including the conventions of the language in use at the time. (Kneale [1972], p. 236.) The language must be specified, because 'a single utterance might conceivably be a specimen of several different propositions according to the conventions of different languages.' (Kneale, [1972], p. 235; *cf.* (*a*), above.) This is given as a leading reason why truth is a property not of utterances or sentences but of propositions. Another reason, which is implied by this one, is that 'for any utterance we choose to consider it is at least conceivable that there should not only be some language in which it would be true . . . but also some language in which it would be false,' so that 'ascription of truth to an utterance without reference to a particular language . . . would not even exclude a like ascription of falsity.' (Kneale [1972], p. 240.)

Now suppose we are given two different languages in which any utterance U we choose to consider expresses distinct propositions, if it expresses any at all. Let one of the languages be English, the other any language L whose conventions are markedly different from those of English. Obviously it is at least conceivable that U does not even express a proposition in L, hence that U does not express a true proposition in L, which is to say that U is not true in L. Thus for any U we choose to consider it is conceivable that U is not true in L. But if for any U it is conceivable that U is not true in L, then, for any U, *that U is not true in* L is a proposition. For 'conceivable that', like 'possible that' and 'true that', makes sense only when what it is prefixed to is a specimen of a proposition. *Therefore an utterance of 'U is not true in L' expresses a proposition no matter what utterance is designated by* '*U*', provided the conventions of *L* are different from those of English. But whatever is true of *all* utterances is true in particular of any utterance designated by 'my present utterance'. (*Cf.* van Fraassen [1970], p. 247.) It follows that my utterance of

(*3*) My present utterance is not true in *L*

expresses a proposition in English, so long as *L* is a language whose conventions are markedly different from those of English. As seen, the conventions of *N* are quite different from those of English. It follows that my utterance of (*1*), 'My present utterance is not true in N', which is abbreviated by (*K*), expresses a proposition in English. *PS*'s presuppose a conception of language which, so far from blocking the assumption that a proposition is expressed, actually implies it.

5 Thus do proposition solutions give rise to a linguistic analogue of the Liar's Revenge, by virtue of their underlying conception of what counts as a fully human language. Naturally those who rely on such solutions will resist this conclusion, but the burden is on them to give a non-*ad hoc* explanation of how and why we may block one or more of the assumptions which jointly yield the inconsistency '$T_N(k)$ and $\sim T_N(k)$'. That is, they must explain, on the basis of features already implicit in their conception of language, why some such assumption is not warranted.

The obvious assumption to question seems to be that there could be a language *N* whose conventions are such that (*i*) any utterance about utterances which is true in English translates into a distinct truth in *N* (*cf. A*, *B*); and (*ii*) an utterance of '$T_N(U)$' translates into the proposition that *U* is not of a sentence of *N* (*cf. C*). Unfortunately there appears to be nothing in *PS*'s which would imply that or explain why no language could have these conventions. On the contrary, that a language could have such conventions is a consequence of the underlying conception of what counts as a fully human language. This follows from the fact that principles (*a*) – (*c*), which are at the heart of that conception, are jointly satisfied by *N*.

Another assumption to question is that my utterance *k* of '$\sim T_N(k)$' expresses a proposition in English. Unfortunately, as lately argued, this assumption also is a consequence of basic features of *PS*'s. There appears to be no plausible way for *PS*'s to avoid the inconsistency.

It would be a mistake to fix the blame simply on the positing of propositions. The antinomy goes through for sentences as well. Even Quine says, 'Conceivably, by an extraordinary coincidence, one and the same string of sounds . . . could serve for '$2<5$' in one language and '$2>5$' in another.' (Quine [1970], p. 14.) Thus where 'S' and 'S^*' represent distinct sentences of English, one and the same utterance U could serve for S in one language and S^* in another (*cf.* (*a*)). Again U expresses a truth only relative to a language L: U is true in L—*i.e.* $T_L(U)$—just if U serves for S in L, and S is true. The sentence S is the truth-bearer.

Now assume what Quine need not assume, that by extraordinary coincidence there is an L in which U serves for 'U is not of a sentence of L'; that in English U serves for '$\sim T_L(U)$'; and that an utterance about utterances which is true in English is true in L. Then U is *not* true in L, since U serves in L for 'U is not of a sentence of L'; and U could not serve for anything in L if U were not of a sentence of L. But in English U serves for '$\sim T_L(U)$', which we now see is true. Therefore U is true in English. Since in English U is about utterances, if $T_E(U)$ then $T_L(U)$. It follows that $T_L(U)$ and $\sim T_L(U)$. This is a linguistic analogue of the Liar's Revenge for sentences. In it the English sentences behave like propositions, as translational constants onto which utterances in different languages are mapped.

It is instructive to see why Quine need not make the additional assumptions which yield the antinomy. Indeed his philosophy would compel us to deny two of them. One is the use of 'language variables'—an uncritical 'dependence on the notion of *a* language.' We must be able to answer the question, 'What are languages, and when do they count as identical or distinct?' (Quine [1960], pp. 213–14.) Failing a satisfactory answer, or a satisfactory principle of individuation for languages, quantifying phrases like 'there is a language L such that . . .', or reference to a language like N which is not actually in use, are out of order.

The other assumption we would be compelled to deny is that we are warranted in *translating* the utterances of sentences of N or of L in the ways specified (*cf.* (*i*)–(*ii*), and A–C). Quine's general thesis of the indeterminacy of translation implies, among other things, that a manual of translation which systematically attributes inconsistent beliefs to the native is suspect for that reason alone. (Quine [1960], p. 59.) The translations we postulated from N to English would require us always to attribute an inconsistent belief to native speakers of N in their utterance k of '$T_N(k)$'. For, as the antinomy shows, their utterance k would be both true and not true in N. Therefore, even if a language like N could somehow be said to exist, in actual practice we would never be warranted in translating it into English in the way required for the antinomy. (Cf. Quine [1970], p. 45, on 'untranslatable foreign language'.)

From this point of view, what is reduced to absurdity is a conception of language according to which translation is determinate, quantification over or reference to possible languages is allowed in the absence of an adequate principle of individuation, and any fully human language enables us to talk self-referentially about talk and about what is said in talk—in that language or at least in some other. Even if defenders of this conception satisfactorily treated Quinian qualms about translation and possible languages, or any other criticisms from outside the pale, they would have to face the trouble from within: internal inconsistency.

Quine's conception of language would provide a non-*ad hoc* explanation of how and why some of the assumptions which yield the antinomy may be blocked. Obviously his is not the only conception which would do so. But by regarding the antinomy initially from a Quinian perspective, we are helped to see that it does not turn simply on the positing of propositions. Sentences, we now know, would do as well. Rather it turns on the notion of what counts as a possible language and what sorts of translations are permitted from one such language to another, regardless of whether translation is construed as the mapping of utterances onto propositions or onto sentences. One may be dissatisfied with ordinary, standard accounts of these matters, or the account presupposed by *PS*'s, without believing in the indeterminacy of translation. The linguistic analogue of the Liar's Revenge helps to show that we ought to be very dissatisfied.

Vanderbilt University

REFERENCES

CHOMSKY, N. [1965]: *Aspects of the Theory of Syntax.*

FRAASSEN, B. C. van [1970]: 'Inference and Self-Reference', *Synthese*, **21**, pp. 425–38.

HERZBERGER, H. [1970]: 'Paradoxes of Grounding in Semantics', *Journal of Philosophy*, **67**, pp. 145–67.

KNEALE, W. [1971]: 'Russell's Paradox and some Others', *British Journal for the Philosophy of Science*, **22**, pp. 321–38.

KNEALE, W. [1972]: 'Propositions and Truth in Natural Languages', *Mind*, **81**, pp. 225–43.

MACKIE, J. L. [1973]: *Truth, Probability and Paradox.*

MARTIN, R. L. [1974]: 'Relative Truth and Semantic Categories: A Reply to J. F. Post', *Journal of Philosophical Logic*, **3**, pp. 149–53.

POST, J. F. [1973]: 'Shades of the Liar', *Journal of Philosophical Logic*, **2**, pp. 370–86.

POST, J. F. [1974]: 'Shades of Possibility: Rejoinder to R. L. Martin', *Journal of Philosophical Logic*, **3**, pp. 155–8.

QUINE, W. V. [1960]: *Word and Object.*

QUINE, W. V. [1966]: *Ways of Paradox.*

QUINE, W. V. [1970]: *Philosophy of Logic.*

RYLE, G. [1951]: 'Heterologicality', *Analysis*, **11**, pp. 61–9.

Brit. J. Phil. Sci. 25 (1974), 235–259 *Printed in Great Britain*

Progress and Degeneration in the 'IQ Debate' (II)

by PETER URBACH

3 THE DEGENERATION OF THE ENVIRONMENTALIST PROGRAMME

(a) *Explanations of IQ Differences in Terms of Socio-Economic and Cultural Factors.*

The hereditarian programme makes no predictions concerning the relative average intelligence of different racial or social groups. It is perfectly consistent with any observed differences (or lack of them) in IQ between different groups. The environmentalist programme, on the other hand, predicts that all groups will be *equal* in innate mental ability; any IQ differences which cannot be explained by environmental factors are consequently anomalies.

In this section I shall discuss some environmental theories which have been used to explain the origins of observed racial and social differences in mean IQ. No doubt many environmentalists will repudiate the idea that some of the environmental theories I discuss were ever part of their programme. But these same environmentalists would surely have claimed victories for their programme had these theories provided progressive explanations for the group differences in IQ.

Environmentalist theories explaining observed group differences in average IQ have consistently lagged behind the facts and auxiliary hypotheses proposed to rescue these theories when their predictions have failed are all *ad hoc*. Occasionally the environmentalist programme has scored some predictive success, but whenever this has happened, hereditarians have caught up and superseded them by reinterpreting the observations within their programme in a non-*ad hoc* way.

Environmentalist attempts to explain the observed differences in average IQ between social classes and between American blacks and whites have sometimes used the auxiliary assumptions (*i*) that the quality of education which an individual receives significantly affects the score he gains on an IQ test and (*ii*) that Negroes and lower class children are more likely to attend schools where facilities are relatively inferior in quality.

The influence of education on a child's IQ was tested by Gordon in his [1923] studies of canal-boat children. These children received no formal education and the environmentalist theory predicted that they would not only have a very low IQ but also that their IQ's would fall as they grew older and lagged further behind in their education. This theory was corroborated. Gordon found that the children had an average Stanford-Binet IQ of only 70 and he also discovered a strong negative correlation between IQ and age.

Hereditarians, however, successfully dealt with this anomaly by the auxiliary assumption that the Stanford-Binet test was 'culture biased' against the canal-boat children because of their extreme cultural isolation and was thus not a true measure of their intelligence. The experiment was repeated using a test of intelligence which required no ability to read or write. And, indeed, the canal-boat children now scored an average IQ of 82. Moreover, there was a slight *positive* correlation between IQ and age.[1]

While educational deprivation may be the cause of low IQ scores in some extreme cases, such as that of the canal-boat children, several investigations have refuted the theory that prevailing differences in school facilities are responsible for the observed overall racial and social differences in IQ scores. In by far the largest investigation of possible influences of the quality of a school's facilities on the intelligence of its pupils, Coleman *et al.* found that for each racial group, most of the variation in IQ scores occurs *within* schools and that less than 20 per cent occurs *between* schools.[2] Moreover, Coleman predicted that 'if these variations [in IQ between schools] were largely a result of either school factors or community differences in support of school achievement, then the school-to-school differences would increase over the grades in school'[3]. This prediction was not corroborated. Coleman found that there was no significant change in school-to-school variation in IQ between grade 1 (when children were tested shortly after they had started school) and grade 6.[4] In another large study of Californian schools which refutes the environmentalist hypothesis, negligible correlations were found between IQ and expenditure per pupil, teachers' salaries, pupil/teacher ratios and the number of school administrators.[5] The results of a further investigation of 900 New York schools showed that, in contrast with environmentalist predictions, there was a strong *negative* correlation between average expenditure per pupil and the average IQ of the pupils.[6]

Halsey has attempted to explain the origin of social class differences in average IQ in terms of the superior cultural and material circumstances enjoyed by the higher classes.[7] But, as Conway points out, Halsey's environmentalist explanation of class differences in IQ is inconsistent with the observation that despite the considerable social and material improvements which took place in the period 1922–50, there was a slight *drop* in average IQ, from 98·3 to 97·1, amongst the children of skilled and unskilled workers in London in this period.[1] In order to deal with this anomaly, Halsey has disputed that the circumstances of the working classes have improved *significantly*. He suggests that 'it is necessary to entertain a sophisticated view of the notion of class environment'[2]. Halsey also goes on to explain that 'this is not simply a matter of crude material differences, but a *complex pattern of attitudes, aspirations, etc.* about which we still know far too little'[3]. The original environmentalist explanation of class differences in terms of *specified* environmental disadvantages has now turned into the 'sophisticated' phrase that these classes have a different and '*complex pattern of attitudes, aspirations, etc.*'[4] This phrase is empirically empty and, indeed, a standard pseudo-scientific manoeuvre of environmentalists. If an hypothesis is contradicted by an anomaly it is intellectually improper to obscure this fact by adding the trivial truth that the world is complex.

---

[1] Gaw [1925], p. 390.

[2] Coleman *et al.* [1966], p. 296.

[3] *Ibid.*, p. 296.

[4] *Ibid.*, pp. 296–7.

[5] Jensen [1971.]

[6] *Cf.* Gittell [1971].

[7] Halsey [1958]. Actually Halsey tried to incorporate an environmentalist account of social class differences into the hereditarian programme. He accepts that up to 75 per cent of the variation in individual IQ's is due to genetic differences and conjectures that the 25 per cent of variation which can be attributed to environmental differences goes largely into causing social class differences.

Donald Swift also regards Conway's contention that the circumstances of the working classes have improved during this century as untenable because it assumes that the environment consists of discrete variables external to the individual. Apparently the 'sociological perspective leads . . . to [the] . . . much more complicated view'[5] that 'the environment of an individual is in his head'[6]. One wonders what empirical observation would contradict this, certainly unusual, theory.

The assumption that the difference in average IQ between American blacks and whites has its origin in environmental factors connected with socio-economic status has also been tested by matching black and white children for socio-economic status and the geographical area of their homes to see whether these groups would have the same average IQ's. The results of such experiments show that groups matched in this way still have large, although somewhat reduced, IQ differences.[7] More significantly, it has been found that black children in the top socio-economic group still performed, on average, slightly less well than children of whites of the lowest socio-economic group.[1] This last observation is an anomaly for the environmentalist programme, because the same observation is easily explained in hereditarian terms if one assumes that both black and white parents in the top socio-economic class have similar average IQ's and that their children regress to different population means.[2]

---

[1] Conway [1959], p. 11.

[2] Halsey [1959], p. 2.

[3] *Ibid.*, p. 2; my italics.

[4] *Ibid.*, my italics.

[5] Swift [1972], p. 154.

[6] *Ibid.*, p. 156.

[7] For a review of these studies *cf.* Shuey, *ibid.*, pp. 518–20. Since IQ and socioeconomic status are positively correlated, groups of children matched for socioeconomic status will also be matched to some extent with respect to their parents' IQ's (*cf. e.g.* Jensen [1972*a*], pp. 235–42). The hereditarian programme therefore *also* predicts that these groups will have smaller average IQ differences than unselected groups, and this result therefore is not a 'severe' test of the rival programmes. (For the concept 'severity of a test', see Popper [1963], p. 388.)

I argued earlier[3] that an important aspect of the heuristics of each of the two research programmes is to try to invent new tests of innate intelligence if a particular test which was originally conjectured to be satisfactory turns out to produce anomalous results. Some environmentalists have carried out this part of their programme. In particular, they conjectured that standard IQ tests, such as the Stanford-Binet test, are 'culture biased' against members of the lower socio-economic classes because 'they do not use problems which are equally familiar and motivating to all such groups'[4] and because they only test a person's 'familiarity with, and training in, middle class linguistic culture'[5]. In an attempt to construct a test not biased in this way, considerable effort was made to select only those types of problem which are 'part of the basic American culture, [and which are] . . . learned relatively as often in all socio-economic and cultural groups'.[6] The suitability of test items was checked by observing children from different social groups in school, at play and in family situations[7] and the problems were couched in symbols which were ascertained, after interviews with children, to have equal familiarity and significance for all social groups.[8] When the resulting 'culture fair' test, the so-called Davis-Eells Games, was used, the social class difference in performance was not significantly different from that found with conventional intelligence tests.[9] Moreover, no other 'culture fair' (that is, non-verbal) test has succeeded in eliminating the class difference in performance.[10]

When culture fair tests are administered to Mexican Indians their scores rise significantly and approach the scores of American whites. This effect is just what is predicted on the assumption that the culture bias of conventional tests prejudices the performance of non-white groups. On the other hand, contrary to the environmentalist prediction, American Negroes do rather *better*, on average, on culture bound than on culture fair tests.[11] Some environmentalists try to account for this anomaly by suggesting that the culture fair tests were not *really* fair.[1] But in order for this assumption to be testable rather than a mere verbal quibble, these environmentalists must specify wherein the bias lies and test their assumption by constructing new tests which are not biased in this way. So far, this has not been done.

---

[1] *Cf.* Shuey, *ibid.*, pp. 519–20, for a review of the relevant studies and Jensen, *ibid.*, pp. 239–40 and pp. 241–2.

[2] *Cf. above*, section 2(*d*) (Part 1), for the same phenomenon exhibited amongst siblings.

[3] *Cf. above*, section 1 (*b*) (Part I).

[4] Davis [1949], p. 48.

[5] *Ibid.*, p. 83.

[6] *Ibid.*, p. 69.

[7] Eells *et al.* [1951].

[8] Davis, *ibid.*, p. 66.

[9] Altus [1956].

[10] *Cf.* MacArthur and Elley [1963]. For a discussion of the Davis Eells and other 'culture fair' tests, *cf. e.g.* Freeman [1963].

[11] *Cf.* Shuey, *ibid.*, pp. 256–8, McGurk [1967], and Coleman [1966], p. 20. Attempts to explain the fact that Negroes, on average, perform better on culture biased than on culture fair tests have contributed to a significant problem shift in the hereditarian

The assumption that social class differences in IQ arise through the beneficial surroundings which higher class parents provide for their children was tested in a well-known experiment carried out by Skodak and Skeels.[2] The result of this experiment is often hailed as a victory for the environmentalist programme, but I shall argue that, on the contrary, when these results are taken as a whole, they *refute* environmentalist predictions.

In the Skodak and Skeels study it was found that the average IQ of some adopted children was 106, while their biological mothers had a mean IQ of only 85·5. Now on the assumption that the fathers of the children had average IQ's, the hereditarian programme anticipates that the childen will have IQ's of about 93. This result is therefore an anomaly for the hereditarian programme. Hereditarians have attempted to deal with the anomaly with several auxiliary assumptions. (*i*) It is conjectured that the fathers' IQ's had been underestimated since they had not been measured and for many of them no information was available even concerning their educational levels.[3] (*ii*) It is suggested that the children were not a random sample of children of low IQ mothers but were specially selected as suitable for adoption.[4] (*iii*) It is conjectured that since the IQ's of the mothers were assessed at an emotionally stressful period around the time of the birth of their child, these may be lower than their true IQ's.[5] None of these conjectures has been independently tested, and they are therefore all *ad hoc*.

Although the results of the Skodak and Skeels study are anomalous for the hereditarian programme, Jensen argues that the IQ bonus acquired by the foster children through their favourable surroundings is simply a repetition of the *old* anomaly that available IQ tests fail to tap more than about 80 per cent of inherited general intelligence. More particularly he argues that if 20 per cent of variation in test scores is caused by environmental variation then an IQ increase close to the one observed is expected since the foster parents 'were at least 1 standard deviation above the general

---

programme. It has been suggested that there are at least two different, although correlated, types of mental ability. Level I is an associative or rote learning ability, while Level II is an abstract reasoning ability. Both types are held to be inherited. Negro children whose scores are much lower than whites on tests of Level II ability, nevertheless score equally well on tests of Level I ability. *Cf.* Jensen [1970*a*].

[1] *Cf. e.g.* Ryan [1972], p. 53.

[2] For the final investigation in this longitudinal study see Skodak and Skeels [1949].

[3] McNemar [1940], pp. 75–6.

[4] Jensen, [1972], p. 17.

[5] McNemar, *ibid.*, p. 75.

average in socio-economic status and were *probably* even higher in other qualities deemed desirable in [such] . . . parents'.[1] But Jensen's explanation is *ad hoc* because it fails to specify the nature of the other desirable qualities which the foster parents 'probably' possessed. Moreover, Jensen's explanation is inconsistent with evidence brought forward by hereditarians concerning the nature of the environmental influences on IQ test results. First, it may be deduced from correlations for IQ of pairs of twins and siblings that, on average, about half of the variation in IQ-relevant environments occurs *within* families.[2] It is therefore unlikely that these IQ determining environmental variations are connected with the cultural and material differences which exist between the social classes. Secondly, after reviewing the evidence that significant IQ determining factors step in during the first stages of cell-division in the development of monozygotic twins, Jensen concludes that 'a substantial and perhaps even a major proportion of the non-genetic variance is attributable to prenatal and other biological influences rather than to differences in the social-psychological environment'.[3] If most of the 20 per cent of the non-genetic contribution to IQ variation arises from biological influences, it is difficult to see how foster parents could make much impression on their adopted children's IQ's.

Thus, *pace* Jensen, the results of the Skodak and Skeels study present the hereditarian programme with a new, unresolved anomaly. However, when the results of the study are taken as a whole, they also disconfirm environmentalist predictions. The prediction that the adopted children would, despite their lower class origins, have IQ's typical of the middle class was based on the assumption that there are major IQ-determining environmental factors correlated with social class. However, this asumption was refuted in the Skodak and Skeels study where there was *no* significant correlation between the occupational class of the adoptive fathers and the IQ's of their foster children. Moreover, there was *no* significant correlation between the children's IQ's and their foster parents' educational levels,[4] a result which has been corroborated in independent studies of children and their foster parents.[5]

Skodak and Skeels recognised that the high IQ's of the adopted children could not be attributed to the educational or occupational levels of the adoptive parents and they concluded that '*other factors*, primarily emotional and personal, and *probably* located in the foster home, *appear* to have more significant influence in determining the mental growth of the child'.[6] Underlining the *ad hoc* nature of this explanation, they add that 'unfortunately, there is still no scale for the measurement of these dynamic aspects of the foster home situation'[1].

---

[1] Jensen [1973*b*], p. 406; my italics.

[2] Jinks and Fulker [1970].

[3] Jensen [1970*b*], p. 144.

[4] Skodak and Skeels, *ibid.*, pp. 108–9.

[5] *Cf. e g.* Honzig [1957] and Burks [1928].

[6] Skodak and Skeels [1949], p. 116; my italics.

As I have already remarked, the hereditarian programme makes no predictions about the relative degrees of intelligence of different racial groups. It takes the IQ values obtained on the 'best' IQ tests as estimates of the true average intelligence levels and proceeds from there to make predictions about persons of known genetic relationships. On the other hand, with the auxiliary assumption that general cultural and socio-economic factors are the most important IQ-determinants, the environmentalist programme predicts that Negroes will, on average, have higher IQ's than American Indians since the latter are disadvantaged relative to the former on a large number of cultural and economic indices frequently considered important by environmentalists.[2] By the same criteria, American orientals should, on average, be less intelligent than whites.

All these environmentalist predictions are controverted by the findings that the order of average IQ ratings of these different groups is: Oriental>White>American Indian>Negro.[3] Thus the comparison of Negroes' and American Indians' IQ's yields a major anomaly for the environmentalist programme. John Rex claims that these results should not be interpreted as anomalous for the environmentalist programme because they fail 'to analyse the structural situation of the American Indian in a reservation and an urban Negro'.[4] But Deutsch's contention that there is a '*greater* degree of structured environmental deprivation within the American Indian community than within the ghetto'[5] if correct, implies that there is no hope for Rex's sub-programme. Deutsch attributes the relatively superior performance of Indians over Negroes to the 'qualitative differences between environments . . . [which] are *probably* highly relevant to any discussion of environment-behaviour relationships'.[6] In particular, Deutsch points to 'the *special* conditions of American Indians: their history, their current social organisation, and their schooling'.[7] This explanation is, in its present form, sufficiently vague to be safe from the dangers of any future test. Another environmentalist, Crow, believes that the interpretation of Negro-Indian differences in IQ as genetic in origin is unacceptable since 'some environmental variable ***associated with being black*** is not included in the environmental rating'.[1] Crow needs only to equate this 'environmental variable associated with being black' with the possession of a lower IQ, to complete the circle.

---

[1] *Ibid.*, p. 116

[2] The group of American Indians which Coleman *et al.* investigated in their [1966] studies were further below Negroes than the Negroes were below whites on the following environmental measures: (1) reading material in the home, (2) cultural amenities in the home, (3) structural integrity of home, (4) foreign language in home, (5) pre-school attendance, (6) encyclopaedia in home, (7) parents' education, (8) time spent on homework, (9) parents' educational desires for child, (10) parents' interest in school work, (11) child's self-concept (self-esteem), (12) child's interest in school and reading.

[3] Coleman *et al.* [1966], p. 20.

[4] Rex [1971].

[5] Deutsch [1969], p. 536; my italics.

[6] *Ibid.*; my italics.

[7] *Ibid.*, p.573; my italics.

The environmentalist assumption that socio-economic and cultural factors play a large part in influencing IQ's is also refuted by the fact that Eskimos living in the Arctic Circle have slightly *higher* mean IQ's than white Canadians and *much higher* average IQ's than American or Jamaican blacks.[2] This result is an anomaly for the environmentalist programme since, as Vernon points out, socio-economic conditions are extremely poor among these Eskimos, there is a similar degree of family instability among Eskimos and Jamaicans and the Eskimos suffer from a high level of unemployment.[3] Sir Peter Medawar suggests that the anomalously high scores gained by Eskimos may be due to the (alleged) fact that 'upbringing in an igloo gives *just the right degree* of cosiness, security and mutual contact to conduce to a good performance in intelligence tests'.[4] Medawar supplies no evidence for this providential balancing of environmental factors; his hypothesis is no more than an empty verbal quibble.

In this section I have reviewed attempts to explain social class and racial differences in IQ in terms of socio-economic and cultural factors. The failure of these attempts to achieve any progress in the environmentalist programme has induced most environmentalists to stop working on social class differences in IQ. They have however continued research into racial differences in IQ, especially those between blacks and whites, and many new and complex factors have been held responsible for these differences. The most important ones are discussed in the following sections. When explaining racial differences in average IQ, the environmentalist programme is sometimes supported by researchers who are hereditarians with regard to differences in intelligence *within* populations. Such researchers justify their apparent change of allegiance by claiming that there are at least two hereditarian IQ programmes, one concerning IQ differences in the white population and the other concerning differences between the black and the white populations and the fact that the first programme is progressive lends no empirical support to the other.

The argument which has been put forward for this position is as follows: it is asserted that the fact that IQ differences are largely innate within the white population is irrelevant when considering the question of whether average IQ differences between the white and black populations reflect environmental or genetic differences. In particular, it is argued that even if all of the IQ variation within the white and black populations is caused by genetic variation, this does not *prove* that differences in IQ between the black and white populations are not the result of environmental differences.[1] Jensen clearly is impressed by this argument. He has replied that 'while . . . heritability within groups cannot *prove* heritability between group means, high within group heritability does increase the *a priori likelihood* that the between groups heritability is greater than zero'.[2] Jensen even offers a mathematical relationship connecting within-group and between-group heritability.[3] Jensen's argument, which rests on the probabilist version of induction, is subject to well-known criticisms.[4] It is trivially the case that any results which are true for one set of individuals are not *necessarily* true for another set. Thus, for example, the degree to which IQ variation is genetic in origin for one 'random' sample of Londoners is not *necessarily* the same as, or even close to, that found in another 'random' sample of the same population. The weakness of the suggestion that black-white IQ differences are entirely environmental in origin even though white-white differences reflect mainly genetic differences is that, without further elaboration, it is *ad hoc*. The suggestion is equivalent to Thoday's claim, which was discussed earlier,[5] that blacks experience a specific IQ-determining environment which affects them all to an equal extent and which is different from the environmental factors which affect the white population. The failure of attempts to corroborate precise versions of this claim will be discussed in the following sections.

---

[1] Crow [1969], p. 308; my italics. Crow's 'environmental variable' is clearly identical with Thoday's *ad hoc* 'environmental factor X'. *Cf. above*, section 2(*d*) (Part I).

[2] MacArthur [1968] and Berry [1966].

[3] Vernon [1965], p. 723.

[4] Medawar [1974]; my italics.

(*b*) *Explanations of Observed IQ Differences in Terms of Personality factors.*

(*bi*) *Motivation, Internal Control, Self-Esteem and Father-Absence.*

A variety of reasons have been suggested why the American Negro should have developed a radically different personality (and hence a radically different average IQ) to his white counterpart. It is said that Negro parents and their children are emotionally and socially isolated because of the emphasis within the Negro family on physical methods of control. This isolation then exposes the children to the excessive social influence of their contemporaries. These factors are said to create a personality marked by feelings of inadequacy and lacking in self-control mechanisms. American Negroes are also said to be lacking in 'achievement motivation' as a result of the matricentric structure of the Negro family and because of the use of child rearing practices which originated in slavery and which were calculated to produce a docile and obedient child.[1]

---

[1] *Cf. e.g.* Lewontin [1970], Bodmer [1972], pp. 105–6 and [1973], p. 13, and Kagan [1969], p. 275.

[2] Jensen [1972], p. 162, footnote; my italics replace those in the original.

[3] *Ibid.*, p. 30, footnote.

[4] For an exposition, *cf.* Lakatos [1968*a*].

[5] *Cf. above*, section 2(*d*) (Part I).

Some researchers have indeed found Negroes to be less strongly motivated towards academic achievement than whites. But these results are unreproducible. Other workers find either no differences between blacks and whites in motivation, or that blacks *exceeded* whites in their desire for achievement.[2]

After reviewing the main theories of deprivation which have been set out to account for the low average IQ of Negroes, Irwin Katz concludes that 'psychologists have contributed little to the understanding of the motivational problems of disadvantaged students' and he adds that 'scientific knowledge has barely advanced beyond the conventional wisdom of the teacher's lounge'.[3]

Katz has conjectured that the two most important determinants of academic success are a sense of ***internal control*** and a high level of ***self-evaluation. Internal control*** refers to the 'degree to which people have a sense of efficacy, or power, and accept personal responsibility for what happens to them'[4]. This quality is held to be more pronounced among white children than among blacks and stronger in the middle class than in the working class. The extent to which individuals possess these characteristics is supposed to affect their expectation of success and hence their willingness to strive. Katz's suggestion is, however, refuted by the results of the 'Upward Bound' programme of educational intervention.[5] These show that while the programme had produced significant increases for both Negroes and whites in measures of self-esteem and internal control, there was no significant change in their school performance. Since achievement at school is highly correlated with IQ, there was presumably no change in the children's IQ's either.

---

[1] For a review of these theories *cf.* Katz [1969].

[2] For a review of the relevant studies, *cf.* Proshansky and Newton [1968], pp. 196–202, and Shuey [1966], p.. 507–8.

[3] Katz, *ibid.*, p. 23. It is surprising that such diffidence about the success of the environmentalist programme should be displayed in one of the major papers in the June 1969 issue of the *Journal of Social Issues* which was explicitly dedicated to the cause of environmentalism. Katz's unassuming claims for environmentalism contrast sharply with the formal statement which opened the same volume in which the council members of the Society for the Psychological Study of Social Issues unanimously claimed that 'there is no direct evidence which supports the view that there is an innate difference between members of different groups . . . [and that] in an examination of Jensen's data, we find that observed racial differences can be attributed to environmental factors'.

[4] Katz, *ibid.*, p. 16.

[5] Incidentally, the results of the 'Upward Bound' programme were reported in the same issue of the *Journal of Social Issues* in which Katz's suggestions are put forward, *cf.* Hunt and Hardt [1969].

Negro children are more frequently brought up in a fatherless home than are white children and some environmentalists of a Freudian inclination hold that the resulting 'excessive' maternal influence leads to an 'imbalance in male and female roles'. This 'personality disorganisation' is said to impair intellectual performance. But whether or not absence of a father in the home causes such an unfortunate personality state or not, the largest and most extensive studies have found no relation at all between this factor and IQ.[1]

(*bii*) *The Sensory Deprivation Theory*

A frequently cited theory which attempts to account for the poorer showing of American blacks in academic achievement is the sensory deprivation hypothesis. According to this theory, the Negro child is often brought up in a home where his sensory experience is severely limited and this deficiency impairs his personality. The degeneration of this particular sub-programme is traced by Baratz and Baratz in their [1970]:

> The first assumption . . . is that the ghetto mother does not provide her child with adequate social and sensory stimulation (Hunt [1961]). However, further research into the ghetto environment has revealed that it is far from a vacuum; in fact there is so much sensory stimulation (at least in the eyes and ears of the middle class researcher) that a contrary thesis was necessarily (*sic*) espoused which argues that the ghetto sensory stimulation is excessive and therefore causes the child to inwardly tune it all out, thus creating a vacuum for himself (Deutsch [1963]). More recently, studies of social interaction suggest that the amount of social stimulation may be quantitatively similar for lower class and middle class children. Thus the *quantitative* deficit explanation now appears, of necessity, to be evolving into a *qualitative* explanation; that is, the child receives as much or even more stimulation as does the middle class child, but the researchers feel this stimulation is not as 'distinctive' for the lower class child as it is for the middle class child (Kagan [1968]).[2]

As the Baratzes point out, even if researchers had found that sensory stimulation was lacking in the homes of poor families this fact would not have corroborated the environmentalist programme since no evidence has ever been provided to support the assumption that sensory deprivation is at all related to intellectual abilities, except in extreme cases of sensory deprivation quite untypical of the ghetto situation.[3]

---

[1] For a review of all the relevant results *cf.* Proshansky and Newton [1968], pp. 202–8. Also, *cf.* Coleman, [1966], pp. 509–11, Jensen [1969], p. 85 and Katz *ibid.*, p. 20.

[2] Baratz and Baratz [1970], p. 37; my italics.

[3] *Cf. ibid.*, p. 37. Nevertheless, this theory has been very influential amongst educationists devising so-called enrichment programmes of compensatory education. *Cf. below*, p. 250.

(*biii*) *The Racial Stress Theory*

According to another theory and one which commands a certain *prima facie* plausibility, the atmosphere of racial tension resulting from the whites' occupying a more prestigious and threatening social position leads to feelings of anxiety in Negroes and this then leads to their impaired performances on intellectual tasks. Katz and his co-workers have attempted to corroborate this theory in a series of ingenious experiments.

Negro students were tested on a digit-symbol substitution test[1] in the presence of two strangers who were either both black or both white. One of the strangers acted as a tester while the other pretended to be a fellow student working on the same task. Stress was imported into the situation by warning the students that they should expect to receive a mild or a severe electric shock during the test.[2]

The students' performances were found to vary in a way depending on the race of the strangers present during the test and the type of electric shock threatened. When a mild shock was threatened, the students performed better on the test in the presence of whites. When they were told they would receive a severe shock, the performance of both groups was depressed, but the threat of a severe shock was more detrimental to performance in the white than in the black condition. This result is consistent with Hull's thesis[3] that mild stress (in the present case, mild shock plus white tester) improves efficiency on mental tasks, while severe stress (in this case, severe shock plus white tester) depresses efficiency.

In a subsequent investigation, Katz attempted to show that an IQ test is equivalent in stress-inducing effect for Negroes to the threat of the *severe* electric shock in the above experiment.

Some male Negro students from a southern college were tested individually—half of them by a Negro and the other half by a white examiner. When the students were told that the test was intended to measure eye-hand coordination, a non-intellectual faculty, they achieved mean scores of 28·9 and 21·4 in the presence of a white and a black examiner, respectively. Later, when the same test was described as an intelligence test, the students gained mean scores of 22·9 and 23·5 when tested by a white and a Negro, respectively.[4]

Describing the test as a test of intelligence apparently stimulated the group which was being examined by a Negro but depressed the group with the white tester. Even granting that these results can be explained in terms of the mild and stimulating stress of the white examiner plus the non-IQ conditions and the severe and depressing stress of a white examiner plus the IQ conditions, it is not clear what light they shed on the differences in blacks' and whites' IQ scores. In neither of the two test conditions did the presence of the white examiner (a putative symbol of the white threat) prove to be a disadvantage for the black students.

---

[1] Digit-symbol substitution tests are sometimes used as sub-tests in conventional intelligence tests.

[2] Katz and Greenbaum [1963].

[3] Hull [1943].

[4] Katz *et al.* [1965]. The differential performances with white and black testers only appeared when the hardest of three tests was used. Katz does not explain why the effect should depend on the difficulty of the task, except in an *ad hoc* way. *Cf.* Katz [1968], p. 276.

Some researchers have claimed that Negroes perform less well on IQ tests when these are administered by a white than when administered by a black but their results are not reproducible. Shuey surveyed nineteen studies of 2,360 southern Negro children in which the examiner was also black and in which the mean IQ was 80·9. In comparable studies of 30,000 southern Negro children in which the examiner was presumed to be white,[1] the mean IQ was 80·6.

Katz has also carried out experiments which were designed to test the conjecture that the stress involved in taking the intelligence test is directly connected with the state of race relations.

In an experiment which Katz *et al.* carried out in 1964,[2] students at a southern black college did better on digit-symbol tests when informed that their scores would be compared with average scores gained by others (that is, other blacks) at their own college than those who were told that their performances would be compared with national averages gained by whites. A control group of white students performed equally well under both national and local-norms conditions. Low motivation cannot be the cause of a lower performance in the national-norms condition because, under this condition, the students rated themselves as caring *more* about doing well than under the local-norms condition. Katz and his co-workers suggest that being compared with white students '*aroused too much drive*' for the black students to work efficiently.[3] However, this explanation has never been independently tested.

Katz advances an alternative explanation which rests on Atkinson's theory that intellectual performance depends on the individual's confidence of success.[4] According to this theory, the greatest efficiency is achieved when there is a 50 per cent subjective probability of success, and both a very low and a very high expectation of success are detrimental to performance. Relying on Atkinson's theory, Katz conjectures that 'the difference between the two test conditions was due to higher subjective probability of success (closer to 0·50) when Negro subjects believed they were competing with members of their own race than when they expected to be compared with whites'.[1]

---

[1] Shuey, *ibid.*, pp. 506–7. In many studies it was not established whether the examiner was black or white. However, if any studies made by black investigators are included in the group where the examiner was presumed to be white, this would, according to the environmentalist theory, mean that 80·6 was an excessive estimate for the average IQ of the 30,000 Negro children in this group. But if this value were revised downward, it would be *a fortiori*, a refutation of the environmentalist assumption that the Negro child is disadvantaged when tested by a white person.

[2] Katz, Epps, and Axelson [1964].

[3] *Ibid.*, p. 83.

[4] Atkinson [1958].

This assumption has been tested and refuted. Katz was able to control the effect of expectancies of success by informing Negro subjects of the statistical probability that they would achieve a score at least average for their age group. The probabilities, which were either 10, 60 or 90 per cent, were purportedly calculated on the basis of a practice trial on a digit-symbol test, but in fact they were arbitrarily assigned. It was found that all students performed more efficiently when the expectation of success was 60 per cent, while performances were about equally poor in both the 10 and 90 per cent expectancy condition, thereby confirming Atkinson's theory. Now Katz predicted that when the subjective probability of success is held constant 'Negro subjects would perform ***better*** . . . when the administrator was white than when he was Negro, or when they were competing with white peers rather than with Negro peers'.[2] But Katz's results show that when the assigned probabilities and the race of the comparison group are kept constant, the subjects performed less well with a white administrator in 50 per cent of the cases.[3] Katz also found, contrary to his prediction, that when the administrator was white, the Negro subjects gained ***higher*** scores when competing with other blacks than when competing with whites. Moreover, when the administrator was black, irrespective of the assigned probabilities and ***even when no probabilities were assigned***, the subjects performed better when competing with whites than with blacks. This last result is the reverse of the one obtained in the experiment carried out in 1964, when the administrator was also black.[4] This anomalous result was ascribed by Katz to regional differences, the earlier experiment (1964) having been done in Florida and the later one (1965) in Tennessee and Katz suggests that '***perhaps*** the Negro student in the Deep South is more fearful of competition with white peers than is the Negro student in the Upper South'.[5]

The results of a further experiment carried out in 1971 also conflicted with those of the experiment of 1964 when it was found that male black students from one northern and one southern college performed

---

[1] Katz [1968], p. 278.

[2] Katz, *ibid.*, p. 278.

[3] *Ibid.*, pp. 279–81.

[4] *Cf. above*, p. 247.

[5] *Ibid.*, p. 281; my italics.

better in the presence of a black examiner when informed that the comparison group was white (white-norms condition) rather than black (black-norms condition).[1]

The conflict between the results of the 1964 and 1971 experiments was explained by assuming that the Negroes regarded white standards as not only a more attractive goal, but also as a more difficult one to achieve. Katz *et al.* suggest that, in the 1964 experiment, 'the depressant motivational effect of having a low expectancy of success (in the white-norm condition) could have outweighed the facilitating influence of the incentive value of success'[2]. In the 1971 experiment, the subjects were of a higher academic standard and it is claimed that they were more confident in their ability to compete intellectually against their white contemporaries.[3]

Even though Katz *et al.* specifically claim that the results of their 1971 experiment were 'as predicted', there is no evidence that this was in fact a prediction. Since Katz has never assigned relative weights to the factors of 'expectancy of success', 'incentive value of success' and academic self-confidence, it is difficult to see how he could have predicted the effect of changing the academic ability of the Negroes he investigated. Moreover, Katz's claim that he predicted the result of his [1971] experiment conflicts with the statement (at the end of the paper in which this result is reported) that '*at present one cannot predict with assurance the specific types of Negro students who may benefit from cross-racial competition* or the specific types of tasks and achievement settings that will reveal behavioural facilitation'.[4]

Furthermore, the original thesis that their subordinate social position is detrimental to Negroes' academic performances now seems to be considerably weakened. For some Negroes at least, social threat and racial stress appear to be of distinct advantage!

Katz's research programme has degenerated. It explained the observed differential intellectual levels of black and white subjects in a *post hoc* way by interpreting them as a result of arbitrary combinations of several plausible factors.[5] (*a*) Feelings of inferiority in the Negro producing a low expectancy of success, (*b*) 'social threat' leading to detrimental emotional reactions and (*c*) 'failure threat' arising from the disapproval of academic failure by 'significant others' may all contribute to a low Negro performance. On the other hand, if these feelings are only mild, then they may produce a *beneficial* response and provide a stimulus for improved performance. An improved performance may also result from the higher motivation associated with the fact that academic success in competition with whites is of 'higher incentive value'. Katz's catch-all matrix of factors has the superficial attraction that it can account for the inferior or superior performance of Negroes under all circumstances. But it also has the disadvantage that it lacks predictive power.[1]

---

[1] Epps, Katz, Perry and Runyon [1971]. [2] *Ibid.*, p. 201. [3] *Ibid.*, p. 201.

[4] *Ibid.*, p. 207; my italics. Katz also admits that his theory has no predictive power in his [1968]. Cf. *below*, p. 250, footnote 1.

[5] The importance of each of these factors is discussed in Katz [1969].

(*c*) *Tests of Environmentalist Theories in Compensatory Education Schemes*

According to the hereditarian programme, differences in intelligence are innate and cannot be altered by social engineering, except when the latter interferes with the action of the genes. On the other hand, the environmentalist programme asserts that those with low IQ's are deprived of one or more material, cultural or psychological factors enjoyed by people with high IQ's. One way in which the environmentalist theories are tested is by trying to increase the IQ scores gained by individuals by supplying them with the putative missing environmental factor or factors. Such tests have been made on a large scale, especially in America, in 'compensatory education' schemes.[2]

Compensatory education schemes are usually devised to test one or more of three theories of deprivation; these are that the low IQ child has insufficient self-esteem,[3] that he suffers from a lack of stimulation at home,[4] and that the low quality of his language inhibits learning.[5]

---

[1] This is admitted by Katz ([1968], p. 271) who says that 'there does not exist at present any comprehensive system of variables for predicting the specific effects of different conditions of stress on the Negro child's performance of various academic tasks'. The *post-hoc* nature of Katz's explanations is particularly clearly illustrated in his juggling with factors when attempting to explain the influence of shock threat on Negroes' intellectual performances:

> 'There are a number of ways of looking at the effects of shock threat. First, if Negro subjects cared more about performing well in the white condition they would have been more fearful lest the strong shock would thus become more salient and distracting. An upward spiral of debilitation could then be set in motion as distraction and fear made the task seem more difficult and this in turn aroused further emotion. Subjects in the Negro environment, on the other hand, had a relatively relaxed attitude toward the task in the low-threat condition (*too* relaxed for good performance). Hence they would not have been fearful of possible decrements due to shock, but *perhaps just enough* concerned to work harder than before. Also relevant to these data is Bovard's . . . notion that the ability to withstand stress is strengthened by the presence of familiar social stimuli that have nurturant associations (in this case other Negroes). (Katz *ibid.*, p. 275; my italics.)

[2] The main task of these schemes was not specifically to raise the IQ scores of the children but to stimulate their general academic performance. That school performance may be changed by environmental manipulation is not ruled out by either the hereditarian or the environmentalist programme.

[3] *Cf. above*, pp. 243–4.

[4] *Cf. above*, p. 245.

[5] The theory that lower class children speak in a 'restricted code' while middle class children acquire an 'elaborated code' of speech and that the intellectual development of lower class children is consequently hampered was proposed by Basil Bernstein. (*Cf.* especially his [1961].) This theory is not supported by any evidence; the large amount of disconfirming evidence is reviewed by Ginsburg in his [1972] (pp. 58–93). Bernstein has now apparently abandoned the theory. (*Cf.* Bernstein [1970])

In the widely publicised 'Headstart' programme in America, a concentrated effort was made in selected deprived areas to provide vastly improved educational facilities as well as to replace inadequate medical and other social facilities. Despite such intensive efforts, Headstart and other programmes have been judged, *from the academic point of view*, to have been a complete failure. IQ's were not raised and school performance was only slightly affected.

In a few small and very intensive programmes, significant increases in some children's IQ's have been claimed. Occasionally these increases are very large.[1] Where unexpected gains in IQ are produced they are accounted for by hereditarians by suggesting that the IQ tests were not culture fair. In particular, Jensen conjectures that the tests were not culture fair when they were first administered because, on this occasion, the children would have found the testing situation unfamiliar and emotionally disturbing.[2] He also conjectures that the compensatory education provided the equivalent of direct coaching for the IQ tests.[3] (It is well-known that coaching for IQ tests produces gains of up to ten points.[4]) These hypotheses are independently testable by repeating the successful compensatory education schemes using culture fair tests and taking precautions to set the children at ease during the test. In fact no compensatory education scheme has yet been reported which raises IQ scores on culture fair tests of intelligence.[5]

The failure of most compensatory education schemes to elicit higher IQ scores has been explained in a variety of ways by environmentalists. According to some, the deficiencies which Negroes suffer, especially in terms of their linguistic and conceptual systems, are too well established by the time the intervention programmes start.[6] Environmentalists have consequently suggested that compensatory education should start when the child is much younger and some have even advocated completely removing children from their homes to be brought up in specialised institutions.[7]

Martin Richards, commenting on the failure of Headstart and other schemes, remarks that 'it is hard to believe that we [the psychologists involved] all could have been so simple minded' and he claims that psychologists have now 'cultivated a feeling for the *complexities of the organisation and structure of the behaviour of children* and the *subtleties of their transactions with the environment*'.[8] When a theory is controverted by an observation, the canons of intellectual honesty demand that the scientist should acknowledge the anomaly and that he should try to account for it in a non-*ad hoc* way. Richards, by not specifying the nature of the subtleties and complexities which he alleges are important, merely plays down the significance of a hard counter-example by a woolly monster-barring strategem.

---

[1] For a review of these programmes, *cf.* Jensen [1969], pp. 104–7.

[2] Jensen [1969], p. 100.

[3] *Ibid.*, pp. 100–1.

[4] Vernon [1954].

[5] *Cf.* Jensen, *ibid.*, p. 101.

[6] *Cf. e.g.* Caldwell [1968].

[7] *Cf. e.g.* Caldwell [1967].

[8] Richards [1973]; my italics.

An interesting conspiracy theory has been suggested to account for the failure of compensatory education. According to this theory, the improved conditions and education which a successful programme would have brought about would have led to the Negroes' gaining a more acute awareness of their social and political position and that this was too great a threat to the power of the established white middle-class. Therefore 'Headstart *had to be* unsuccessful because success would have led to fundamental social and political changes that the non-poor were not prepared to accept'.[1] Such untestable hypotheses have no place within science.

(*d*) *The Inheritance of the Environment*

Environmentalists, in bending over backwards to preserve their theory intact, have sometimes inadvertently, toppled over into the hereditarian camp. The biologist Steven Rose, for instance, claims that 'nutritional-deficiency in childhood results in permanent changes in brain chemistry for which a subsequent adequate diet does not compensate.[2] Indeed, there is some evidence relating early nutritional deficiencies in children to their later IQ scores and subsequent researches along these lines may well make significant contributions to the environmentalist programme. But Rose goes further and claims that 'such effects are transmitted between one generation and the next', and furthermore that 'brain structure and chemistry [which] determine performance are subtly but profoundly affected by *immediate* environmental factors and by those stretching back beyond our own generation *into an indeterminate distance into the past*'.[3] But if the time when the important environmental factors step in is '*indeterminate*', then Professor Rose's theory cannot be tested and it is scientifically valueless.[4]

It is held further that not only environmentally induced chemical factors may be inherited but that there are some 'cultural patterns that are so resistant to alteration that they have the *appearance* of being innate; indeed, the difficulties in changing the attitudes to school performance and in changing IQ in deprived populations reflect, in part, the difficulty in changing a cultural pattern.'[1] John Rex considers that 'the *crucial* variable is the difference between white and Negro history and the fact that Negro history involves the fact of slavery'[2]. Consequently any experiment designed to compensate for the differences in environment between black and white 'would mean subjecting the group of Negroes to white experience over several hundred years, or subjecting a group of whites to Negro experience'[3]. Rex admits that no experiment can be designed to test his assumption but he regards this as 'fortunate'[4] for the hereditarian programme. But Popper surely taught us that while such facts may be seen as 'fortunate' by a pseudo-scientist, they must be regarded as *unfortunate* by the scientist.

---

[1] Richards, Richardson and Spears [1972], p. 184; my italics.

[2] Rose [1972], p. 143.

[3] *Ibid.*, p. 143; my italics.

[4] It is interesting that there is considerable similarity between Rose's theory and the hereditarian theory proposed by Galton; both assert that intelligence is inherited, in a non-Mendelian manner.

(*e*) *Conclusion*

In this Part of my paper I have considered some of the predictions made by the environmentalist programme, especially in regard to social class and racial differences in average IQ. I have shown that almost none of these predictions has been confirmed and that when predictions have failed, environmentalists have rescued their theories in an *ad hoc* fashion. This patching-up process has left the environmentalist programme as little more than a collection of untestable theories which provide a '*passe partout* which explains everything because it explains nothing'[5]. For example, Bodmer has concluded that the differences in average IQ between American blacks and whites 'could be explained by environmental factors, many of which we still know nothing about'[6]. Professor Bodmer is of course right: everything in the world can be explained by factors about which we know nothing.

The fact that the environmentalist programme has been degenerating does not mean that no progressive programme will ever be based on its hard core. Of the infinite number of possible IQ-determining environmental factors only a handful has been investigated. Resourceful environmentalists of the future may well invent a powerful heuristic which will lead them to *content-increasing* explanations of individual and group differences in IQ's.[1]

---

[1] Bodmer [1972], p. 87; my italics.

[2] Rex [1972], p. 170; my italics.

[3] Rex, *ibid.*, p. 170.

[4] *Ibid.*, p. 87.

[5] This perceptive phrase was attributed by Isaiah Berlin ([1939], p. 118) to Marx. It appears, however, that the credit should go to Berlin himself, since he apparently interpolates the phrase into a translation of Marx's celebrated letter to Mikhailovski (Marx [1877]). Imre Lakatos called my attention to the fact that Berlin's deep *aperçu* was later repeated (unfortunately without acknowledgment) by two other great thinkers of this century in almost the very same year. In Agatha Christie's words: 'Dr. Maverick [a psychologist] can explain *anything* [therefore it is not worth listening]'. (Christie [1952], p. 144.) Karl Popper, a year later, put it still better, when he referred to 'those impressive and all-explanatory theories which act upon weak minds like revelations'. (Popper [1963], p. 39.)

[6] Bodmer [1972], p. 112.

REFERENCES

ALTUS, G. T. [1956]: 'Some correlates of the Davis-Eells tests', *Journal of Consulting Psychology*, **20**, pp. 227–32.

ANASTASI, A. [1958]: *Differential Psychology*, Third edition.

ATKINSON, J. W. [1958]: 'Towards experimental analysis of human motives in terms of motives, expectancies, and incentives', in J. W. Atkinson (*ed.*): *Motives in fantasy, action, and society*, pp. 288–305.

BAJEMA, C. J. [1963]: 'Estimation of the direction and intensity of natural selection in relation to human intelligence by means of the intrinsic rate of natural increase', *Eugenics Quarterly*, **10**, pp. 175–87. Reprinted in H. J. Butcher and D. E. Lomax (*eds.*): *Readings in Human Intelligence*, pp. 249–264.

BARATZ, S. S. and BARATZ, J. C. [1970]: 'Early childhood intervention: the social science base of institutional racism', *Harvard Educational Review*, **40**, pp. 29–50.

BENDIX, R. and LIPSET, S. M. [1960]: *Social Mobility in Industrial Society*.

BERLIN, I. [1939]: *Karl Marx*.

BERNSTEIN, B. [1961]: 'Social class and linguistic development: a theory of social learning', in A. H. Halsey, J. Floud and C. A. Anderson (*eds.*): *Education, Economy, and Society*, pp. 288–314.

BERNSTEIN, B. [1970]: 'Education cannot compensate for society', *New Society*, No. 387, pp. 344–7.

BERRY, J. W. [1967]: 'Temne and Eskimo perceptual skills', *International Journal of Psychology*, **1**, pp. 207–22.

BINET, A. [1911]: 'Nouvelles recherches sur la mésure du niveau intellectuel chez les enfants d'ecole', *L'Année Psychologique*, **17**, pp. 145–201.

BINET, A. and HENRI, V. [1895]: 'La psychologie individuelle', *L'Année Psychologique*, **2**, pp. 411–65.

BINET, A. and SIMON, T. [1905*a*]: 'Methodes nouvelles pour le diagnostic du niveau intellectuel des anormaux', *L'Année Psychologique*, **11**, pp. 191–244.

BINET, A. and SIMON, T. [1905*b*]: 'Applications des methodes nouvelles au diagnostic du niveau intellectuel chez des enfants normaux et anormaux d'hospice et d'ecole primaire', *L'Année Psychologique*, **11**, pp. 245–336.

BODMER, W. F. [1972]: 'Race and IQ: the genetic background', in K. Richardson, D. Spears and M. Richards (*eds.*): *Race, Culture and Intelligence*, pp. 83–113.

BODMER, W. F. [1973]: 'Jensenism', *The Listener*, **89**, 4 January, pp. 12–16.

BRUSH, S. G. [1965]: *Kinetic Theory*, **1**.

BURKS, B. S. [1928]: 'The relative influence of nature and nurture upon mental development: a comparative study of foster parent–foster child resemblance and true parent-true child resemblance', *27th Yearbook of the National Society for the Study of Education*, **27**, pp. 219–316.

BURT, C. [1917]: *The Distribution and Relations of Educational Abilities* (London County Council Reports, no. 1868).

BURT, C. [1935]: *The Subnormal Mind*.

BURT, C. [1943]: 'Ability and income', *The British Journal of Educational Psychology*, **13**, pp. 83–98.

---

[1] Neither Bodmer nor Cavalli-Sforza is likely to be among this group of tenacious environmentalists. These two scientists wish to discourage research into the field of racial differences in intelligence because, at present, there are obstacles to *proving* whether these differences are genetically determined and because such studies might be 'misinterpreted as a form of racism' (in their otherwise excellent [1971], p. 800). (Professors Bodmer and Cavalli-Sforza might like to be reminded of the days when Relativity theory was 'misinterpreted' as a form of Semitism.)

The question whether the search for truth should be abandoned out of deference to political sensitivities is not discussed in this paper.

Burt, C. [1955]: 'The evidence for the concept of intelligence', *The British Journal of Educational Psychology*, **25**, pp. 158–77.

Burt, C. [1957]: 'The distribution of intelligence', *The British Journal of Psychology*, **48**, pp. 161–75.

Burt, C. [1959]: 'Class differences in general intelligence: III', *The British Journal of Statistical Psychology*, **12**, pp. 15–33.

Burt, C. [1961*a*]: 'Intelligence and social mobility', *The British Journal of Statistical Psychology*, **10**, pp. 33–63.

Burt, C. [1961*b*]: 'The gifted child', *The British Journal of Statistical Psychology*, **14**, pp. 123–39.

Burt, C. [1963]: 'Is intelligence distributed normally?', *The British Journal of Statistical Psychology*, **16**, pp. 175–90.

Burt, C. [1966]: 'The genetic determination of differences in intelligence: a study of monozygotic twins reared together and apart', *The British Journal of Psychology*, **57**, pp. 137–53.

Burt, C. [1971]: 'Quantitative genetics in psychology', *The British Journal of Mathematical and Statistical Psychology*, **24**, pp. 1–21.

Burt, C. and Howard, M. [1956]: 'The multifactorial theory of inheritance and its application to intelligence', *The British Journal of Statistical Psychology*, **9**, pp. 95–131.

Burt, C. and Howard, M. [1957]: 'Heredity and intelligence: a reply to criticisms', *The British Journal of Statistical Psychology*, **10**, pp. 33–63.

Caldwell, B. [1967]: 'What is the optimal learning environment for the young child?', *American Journal of Orthopsychiatry*, **37**, pp. 9–21.

Caldwell, B. [1968]: 'The fourth dimension in early childhood education' in R. Hess and R. Bear (*eds.*): *Early Education: Current Theory, Research and Action.*

Carter, C. O. [1966]: 'Differential fertility by intelligence', in J. E. Meade and A. S. Parkes (*eds.*): *Genetic and Environmental Factors in Human Ability*, pp. 185–200.

Cattell, J. McK. [1890]: 'Mental tests and measurements', *Mind*, **15**, pp. 373–81.

Cattell, R. B. [1937]: *The Fight for our National Intelligence.*

Cattell, R. B. [1950]: 'The fate of national intelligence: test of a thirteen-year prediction', *The Eugenics Review*, **42**, pp. 136–48.

Cavalli-Sforza, L. L. and Bodmer, W. F. [1971]: *The Genetics of Human Populations.*

Christie, A. [1952]: *They do it with mirrors.*

Coleman, J. S. *et al.* [1966]: *Equality of Educational Opportunity.*

Conway, J. [1958]: 'The inheritance of intelligence and its social implications', *The British Journal of Statistical Psychology*, **11**, pp. 171–90.

Conway, J. [1959]: 'Class differences in general intelligence II', *The British Journal of Statistical Psychology*, **12**, pp. 5–14.

Crow, J. F. [1969]: 'Genetic theories and influences: Comments on the value of diversity', *Harvard Educational Review*, **39**, pp. 301–9.

Daniels, J. and Houghton, V. [1972]: 'Jensen, Eysenck and the eclipse of the Galton Paradigm', in K. Richardson, D. Spears and M. Richards (*eds.*): *Race, Culture and Intelligence*, pp. 68–80.

Davis, A. [1949]: *Social Class Influences upon Learning.*

Descartes, R. [1637]: *Discourse on Method*, in E. S. Haldane and G. R. T. Ross (*eds.*) *The Philosophical Works of Descartes.*

Deutsch, C. P. [1963]: 'Auditory discrimination and learning: social factors', *Merrill-Palmer Quarterly of Behavior and Development*, **10**, (3), pp. 277–96, reprinted in M. Deutsch (*ed.*): *The Disadvantaged Child*, 1967.

Deutsch, M. [1969]: 'Happenings on the way back to the forum: social science, IQ, and race differences revisited', *Harvard Educational Review*, **39**, pp. 523–54.

Eells, K., Davis, A., Havighurst, R. J., Herrick, V. E. and Tyler, R. [1951]: *Intelligence and Cultural Differences.*

Epps, E. G., Katz, I., Perry, A. and Runyon, E. [1971]: 'Effect of race of comparison referent and motives on Negro cognitive performance', *Journal of Educational Psychology*, **62**, pp. 201–8.

Eysenck, H. J. [1939]: "Review of L. L. Thurstone's 'Primary mental abilities'", *The British Journal of Educational Psychology*, **9**, pp. 270–5.

Eysenck, H. J. [1971*a*]: *Race, Intelligence and Education.*

Eysenck, H. J. [1971*b*]: Letter in *New Society*, 8 July.

Eysenck, J. J. [1973]: *The Inequality of Man.*

Falconer, D. S. [1960]: *Introduction to Quantitive Genetics.*

Freeman, F. N. [1939]: *Mental Tests.*

Freeman, F. S. [1963]: *Theory and Practice of Psychological Testing.*

Fisher, R. A. [1918]: 'The correlation between relatives on the supposition of Mendelian inheritance', *Transactions of the Royal Society of Edinburgh*, **52**, pp. 399–433.

Galton, F. [1865]: 'Hereditary talent and character', *Macmillan's Magazine*, **12**, pp. 157–66, 318–27.

Galton, F. [1869]: *Hereditary Genius: an enquiry into its laws and consequences.*

Galton, F. [1883]: *Inquiries into Human Faculty and its Development.*

Galton, F. [1889]: *Natural Inheritance.*

Gaw, F. [1925]: 'A study of performance tests', *The British Journal of Psychology*, **15**, pp. 374–92.

Ginsburg, H. [1972]: *The Myth of the Deprived Child.*

Gittell, M. [1971]: *New York City School Fact Book.*

Glass, D. V. [1954]: *Social Mobility in Britain.*

Gordon, H. [1923]: *Mental and Scholastic Tests among Retarded Children*, Board of Education, Education Pamphlet No. 44.

Halsey, A. H. [1958]: 'Genetics, social class, and intelligence', *British Journal of Sociology*, **9**, pp. 15–28.

Halsey, A. H. [1959]: 'Class differences in general intelligence: I', *The British Journal of Statistical Psychology*, **12**, pp. 1–4.

Herrnstein, R. J. [1971]: 'I.Q.', *The Atlantic Monthly*, pp. 43–64.

Herrnstein, R. J. [1973]: *IQ in the Meritocracy.*

Higgins, J., Reed, E. and Reed, S. [1962]: 'Intelligence and family size: a paradox resolved', *Eugenics Quarterly*, **9**, pp. 84–90.

Honzig, M. P. [1957]: 'Developmental studies of parent-child resemblance in intelligence', *Child Development*, **28**, pp. 125–8.

Hudsen, L. [1971]: 'Science and Popularisation', *New Society*, pp. 29–30.

Hull, C. L. [1943]: *Principles of Behavior.*

Hunt, J. McV.: [1961]: *Intelligence and Experience.*

Hunt, D. E. and Hardt, R. H. [1969]: 'The effect of Upward Bound programmes on the attitudes, motivation and academic achievement of Negro students', *Journal of Social Issues*, **25**, pp. 117–29.

Huntley, R. M. C. [1966]: 'Heritability of Intelligence', in J. E. Meade and A. S Parkes (*eds.*): *Genetic and Environmental Factors in Human Ability*, pp. 201–18.

Jencks, C. [1972]: *Inequality.*

Jensen, A. R. [1969]: 'How much can we boost IQ and scholastic achievement?', *Harvard Educational Review*, **39**, pp. 1–123.

Jensen, A. R. [1970*a*]: 'A theory of primary and secondary familial mental retardation' in N. R. Ellis (*ed.*) *International Review of Research in Mental Retardation*, **4**, pp. 33–105.

Jensen, A. R. [1970*b*]: 'IQ's of identical twins reared apart', *Behaviour Genetics*, **1**, pp. 133–48.

Jensen, A. R. [1971]: 'Do schools cheat minority children?', *Educational Research*, **14**, pp. 3–28.

Jensen, A. R. [1972]: *Genetics and Education.*

Jensen, A. R. [1973*a*]: *Educability and Group Differences.*

Jensen, A. R. [1973*b*]: *Educational Differences.*

Jinks, J. L. and Fulker, D. W. [1970]: 'Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behaviour', *Psychological Bulletin*, **73**, pp. 311–49.

Kagan, J. C. [1968]: 'His struggle for identity', *Saturday Review*, December, 1968.

Kagan, J. S. [1969]: 'Inadequate evidence and illogical conclusions', *Harvard Education Review*, 39, pp. 274–7.

Katz, I., [1968]: 'Factors influencing Negro performance in the desegregated school', in M. Deutsch, I. Katz and A. R. Jensen (*eds.*): *Social Class, Race and Psychological Development*, pp. 254–89.

Katz, I. [1969]: 'A critique of personality approaches to Negro performance, with research suggestions', *Journal of Social Issues*, **25**, pp. 13–27.

Katz, I. and Greenbaum, C. [1963]: 'Effects of anxiety, threat and racial environment on task performance of Negro college students', *Journal of Abnormal Social Psychology*, **66**, pp. 562–7.

Katz, I., Epps, E. G. and Axelson, L. J. [1964]: 'Effect upon Negro digit-symbol performance of anticipated comparison with whites and with other Negroes', *Journal of Abnormal and Social Psychology*, **69**, pp. 77–83.

Katz, I., Roberts, S. O. and Robinson, J. M. [1965]: 'Effects of difficulty, race of administrator, and instructions on Negro digit-symbol performance', *Journal of Personality and Social Psychology*, **70**, pp. 53–9.

Kelley, T. L. [1928]: *Crossroads in the mind of man: A study of differentiable mental abilities.*

Lakatos, I. [1963–1964]: 'Proofs and refutations', *The British Journal for the Philosophy of Science*, **14**, pp. 1–25, 120–39, 221–43, 296–342.

Lakatos, I. [1968*a*]: 'Changes in the Problem of Inductive Logic', in I. Lakatos (*ed.*): *The Problem of Inductive Logic*, pp. 315–417.

Lakatos, I. [1968*b*]: 'Criticism and the Methodology of Scientific Research Programmes', *Proceedings of the Aristotelian Society*, **69**, pp. 149–86.

Lakatos, I. [1970]: 'Falsification and the Methodology of Scientific Research Programmes' in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge*, pp. 91–195.

Lakatos, I. [1971*a*]: 'History of Science and its Rational Reconstructions' in R. C. Buck and R. S. Cohen (*eds.*): *Boston Studies in the Philosophy of Science*, **8**, pp. 91–136.

Lakatos, I. [1971*b*]: 'Popper zum Abgrenzungs- und Induktionsproblem' in H. Lenk (*ed.*): *Neue Aspekte der Wissenschaftstheorie*, pp. 75–110; translated into English as 'Popper on Demarcation and Induction' in P. A. Schilpp (*ed.*): *The Philosophy of Sir Karl Popper*, 1974.

Lakatos, I. and Zahar, E. G. [1975]: 'Why did Copernicus's Programme supersede Ptolemy's?', in R. S. Westman (*ed.*): *The Copernican Achievement.*

LeRoy [1899]: 'Science et Philosophy', *Revue de Métaphysique et de Morale*, **9**, pp. 503–62, 706–31.

Lewis, D. G. [1957]: 'The normal distribution of intelligence: a critique', *The British Journal of Psychology*, **48**, pp. 98–104.

Lewontin, R. C. [1970]: 'Race and Intelligence', *Bulletin of the Atomic Scientists*, **26**, pp. 2–8.

Lipset; S. M. and Zetterberg, H. L. [1956]: 'A theory of social mobility', *Transactions of the Third World Congress of Sociology*, **3**, pp. 155–77.

MacArthur, R. S. and Elley, W. B. [1963]: 'The reduction of socio-economic bias in intelligence testing', *The British Journal of Educational Psychology*, **33**, pp. 107–19.

MacArthur, R. S. [1968]: 'Some differential abilities of northern Canadian native youth', *International Journal of Psychology*, **3**, pp. 43–51.

McGurk, F. C. J. [1967]: 'The culture hypothesis and psychological tests', in R. E. Kuttner (*ed.*): *Race and Modern Science.*

McNemar, Q. [1940]: 'A critical examination of the University of Iowa studies of environmental influences upon the IQ', *Psychological Bulletin*, **37**, pp. 63–92.

McNemar, Q. [1942]: *The Revision of the Stanford-Binet Scale.*

Maddox, H. [1957]: 'Nature-nurture balance sheets', *The British Journal of Educational Psychology*, **27**, pp. 166–75.

Marx, K. [1877]: 'Letter to Mikhailovski' in K. Marx and F. Engels [1953].

Marx, K. and Engels, F. [1953]: *Ausgewählte Briefe*, Berlin Dietz.

Medawar, P. E. [1974]: 'More unequal than others', *New Statesman*, 11 January.

Milhaud [1896]: 'La Science Rationelle', *Revue de Métaphysique et de Morale*, **4**, pp. 280–302.

Morant, G. M. [1956]: 'The significance of racial differences', in *The Race Question in Modern Science*, pp. 285–321, UNESCO.

Miller, S. M. [1960]: 'Comparative social mobility', *Current Sociology*, **9**, pp. 1–89.

Newman, H. H., Freeman, F. N. and Holzinger, K. J. [1937]: *Twins: A study of heredity and environment.*

PEARSON, K. [1904]: 'On the laws of inheritance of man II', *Biometrica*, **3**, pp. 131–90.

PENROSE, L. [1963]: *The Biology of Mental Defect*, 3rd edition.

POPPER, K. [1934]: *Logik der Forschung*, 1935 (Expanded English edition: Popper [1959]).

POPPER, K. [1959]: *The Logic of Scientific Discovery.*

POPPER, K. [1963]: *Conjectures and Refutations.*

POPPER, K. [1972]: *Objective Knowledge.*

PROSHANSKY, H., and NEWTON, P. [1968]: 'The nature and meaning of Negro self-identity', in M. Deutsch, I. Katz and A. R. Jensen (*eds.*): *Social Class, Race, and Psychological Development*, pp. 178–218.

REX, J. [1971]: Letter in *New Society*, No. 456.

REX, J. [1972]: 'Nature versus nurture: the significance of the revived debate', in M. Richards, K. Richardson and D. Spears (*eds.*): *Race, Culture and Intelligence*, pp. 167–78.

RICHARDS, M. [1973]: 'Putting Jensenism in its proper place', *The Times Higher Education Supplement*, 20th July, p. 11.

RICHARDS, M., RICHARDSON, K. and SPEARS, D. [1972]: 'Conclusions: intelligence and society', in M. Richards, K. Richardson and D. Spears (*eds.*): *Race, Culture and Intelligence*, pp. 179–96.

RICHARDSON, K. and HOUGHTON, D. [1973]: 'Race and IQ—a reply', *New Humanist*, **89**, pp. 159–60.

RICHMOND, W. K. [1953]: 'Educational measurement: its scope and limitations: a critique', *The British Journal of Psychology*, **44**, pp. 221–31.

ROBERTS, J. A. F. [1952]: 'The genetics of mental deficiency', *Eugenics Review*, **44**, pp. 71–83.

ROSE, S. [1972]: 'Environmental effects on brain and behaviour', in M. Richards, K. Richardson and D. Spears (*eds.*): *Race, Culture and Intelligence*, pp. 128–144.

RYAN, J. [1972]: 'I.Q.—the illusion of objectivity', in K. Richardson, D. Spears and M. Richards (*eds.*): *Race, Culture and Intelligence*, pp. 36–55.

SHUEY, A. M. [1966]: *The Testing of Negro Intelligence*, Second edition.

SIMON, B. [1971]: *Intelligence, Psychology and Education, a Marxist Critique.*

SKODAK, M. and SKEELS, H. M. [1949]: 'A final follow-up study of one hundred adopted children', *Journal of Genetic Psychology*, **75**, pp. 85–125.

SPEARMAN, C. [1904]: 'General intelligence objectively determined and measured', *American Journal of Psychology*, **115**, pp. 210–92.

SPEARMAN, C. [1927]: *The Abilities of Man.*

SPEARMAN, C. [1939]: 'Thurstone's work reworked', *The Journal of Educational Psychology*, **30**, pp. 1–16.

SWIFT, D. [1972]: 'What is the Environment?', in K. Richardson, D. Spears and M. Richards (*eds.*): *Race, Culture and Intelligence*, pp. 147–66.

TERMAN, L. M. [1916]: *The Measurement of Intelligence.*

TERMAN, L. M. and MERRILL, M. A. [1937]: *Measuring Intelligence.*

TERMAN, L. M. and MERRILL, M. A. [1960]: *Stanford-Binet Intelligence Scale.*

THODAY, J. M. [1973]: 'Educability and Group Differences', *Nature*, **245**, pp. 418–20.

THOMPSON, G. H. *et al.* [1949]: *The Trend of Scottish Intelligence.*

THORNDIKE, E. L. [1914]: *Educational Psychology.*

THORNDIKE, E. L. [1925]: *The Measurement of Intelligence.*

THURSTONE, L. L. [1938]: *Primary Mental Abilities*, Monographs of the Psychometric Society, **1**.

THURSTONE, L. L. [1948]: 'Psychological implications of factor analysis', *American Psychologist*, **3**, pp. 402–8.

THURSTONE, L. L. and THURSTONE, T. G. [1941]: *Factorial Studies of Intelligence.*

U.N.E.S.C.O. [1950]: 'Statement on Race', in *What is Race?*, U.N.E.S.C.O., 1952, pp. 76–80.

VERNON, P. E. [1950]: *The Structure of Human Abilities.*

VERNON, P. E. [1951]: 'Recent investigations of intelligence and its measurement', *Eugenics Review*, **43**, pp. 125–37.

VERNON, P. E. [1954]: 'Symposium on the effects of coaching and practice in intelligence tests', *The British Journal of Educational Psychology*, **35**, pp. 5–8.

VERNON, P. E. [1965]: 'Ability factors and environmental influences', *American Psychologist*, **20**, pp. 723–33.

WARBURTON, F. W. [1969]: 'The British Intelligence Scale', in B. Dockrell (*ed.*): *On Intelligence*, pp. 71–98.

WATKINS, J. W. N. [1958]: 'Confirmable and influential metaphysics', *Mind*, **67**, pp. 344–65.

WATSON, J. B. [1931]: *Behaviorism.*

WECHSLER, D. [1944]: *The Measurement of Adult Intelligence.*

WISSLER, C. [1901]: 'The correlation of mental and physical tests', *Psychological Review Monograph*, **3**, pp. 1–63.

WOLF, T. H. [1969]: 'The Emergence of Binet's Conceptions and Measurement of Intelligence: A Case History of the Creative Process', *Journal of the History of the Behavioural Sciences*, **5**, pp. 113–34, 207–37.

YERKES, R. M. [1921]: *Psychological Examining in the United States Army.*

ZAHAR, E. [1973]: 'Why did Einstein's programme supersede Lorentz's?', *The British Journal for the Philosophy of Science*, **24**, pp. 95–123, 223–62.

Brit. J. Phil. Sci. **25** (1974), 261–289 *Printed in Great Britain*

Discussions

ON THE INFIRMITIES OF GILLIES'S RULE

In his [1971], Donald Gillies suggested that no adequate rule for regarding probabilistic and statistical hypotheses as 'falsified' had hitherto been formulated and attempted to supply a rule that

> will not merely be a description of good scientific practice, but rather a normative proposal that probability should be dealt with in accordance with its dictates (p. 232).

I believe that Gillies's rule is neither a good description of good scientific practice, nor has any value as a normative proposal. For (1) it pervasively sanctions contradictory inferences, and (2) it is widely at variance with established practice for one of the most common types of problems that arise in research.

The claim made in (1) is a development of the points made by De Maré in his comment on Gillies's paper.[1] Let H be of the form 'Random variable Z defined over experiment E has density function f', where f is continuous over the range of Z. Let W be the random variable which takes on the value $\int_{-\infty}^{z} f(t)dt = F(z)$ when Z takes on the value z. As De Mare points out, W is uniformly distributed over the interval (0,1). Let X be the random variable *tan* $(\pi\,(W - \frac{1}{2}))$. The range of X is the real line and X is Cauchy distributed. This follows from the fact that X is an increasing function of W, and the fact that the derivative of W with respect to x equals $(\pi\,(1 + x^2)\,)^{-1}$. For, letting G be the distribution function of X, $G(x) = \int_{0}^{w} dw = \int_{-\infty}^{x} (dW/dx)\, dx$. The Cauchy distribution is unimodal, symmetrical about its mode (which occurs when $X = 0$), and is highly falsifiable under Gillies's Rule. Let H' denote the hypothesis that X has the density $(\pi\,(1 + x^2))^{-1}$ over the real line. H' is, of course, equivalent to H. Gillies would recommend rejecting H' if and only if X takes on a value in the class $\{x \mid x < -x_r \text{ or } x > x_r\}$ where x_r is such that the k and l values for this class are suitable. If $x_r = 12.35$, $k = .05$ and $l = .007$. According to Gillies's remarks on *p*. 240, this is a highly satisfactory rejection class.

Now consider the random variable Y defined over the possible outcomes of E as follows: $Y =$ *cotan* $(\pi(W - \frac{1}{2})\,)$. Y is a decreasing function of W over the interval $(0,\frac{1}{2})$ and also over the interval $(\frac{1}{2},1)$. The range of Y is the real line minus zero. $dW/dy = -(\pi(1 + y^2)\,)^{-1}$. The density distribution of Y is obtained as follows: Let G be the cumulative distribution function of Y and let y be less than zero. $G(y) = \int_{w}^{\frac{1}{2}} dw = \int_{y}^{-\infty} (dW/dy)\, dy = -\int_{-\infty}^{y} (dW/dy)\, dy = \int_{-\infty}^{y} (\pi(1 + y^2)\,)^{-1}\, dy$. For $y > 0$, $G(y) = \frac{1}{2} + \int_{w}^{1} dw = \frac{1}{2} + \int_{y}^{0} (dW/dy)dy = \frac{1}{2} + \int_{0}^{y} (\pi\,(1 + y^2)\,)^{-1}\, dy = \int_{-\infty}^{y} (\pi\,(1 + y^2)\,)^{-1}\, dy$. Thus Y has exactly the same density distribution as X. Let H'' stand for the hypothesis that Y has this distribution (the unit

---

[1] *Cf.* De Maré [1972].

Cauchy). The falsification class for H'' has the same form as the one for H', *i.e.* it is the class $\{y \mid y < -12.35 \text{ or } y > 12.35\}$. But the class for rejecting H' corresponds to the set of outcomes of E for which Z takes on a value $> F^{-1}(.975)$ or $< F^{-1}(.025)$, whereas the class for rejecting H'' corresponds to the class of outcomes for which $F^{-1}(.475) < Z < F^{-1}(.525)$. However, H is equivalent to both H' and H''. Thus Gillies's rule recommends two non-overlapping and quite different falsification classes for H. Call the class based on H', C' and that based on H'', C''. The above argument applies to the vast majority of testing problems encountered in scientific research, for the hypothesis of interest is usually tested indirectly by means of hypotheses like H.

The conclusion to be drawn from the above is that either the tests corresponding to C' and C'' are both valid or Gillies's rule is inadequate. Gillies's discussion of an example of Neyman's in section 4 of his paper indicates that he finds no difficulty in declaring both tests to be valid. Neyman's example embodies a concrete illustration of the type of argument developed above. It presents an hypothesis which implies that certain random variables X and Y have the same distribution, where X and Y are such that $|XY| < 1$. The test generated by X on Gillies's rule is like C' and that generated by Y is like C''. Gillies claims that they are both valid and that there is nothing paradoxical about this. His argument to establish the non-paradoxicality of the existence of such tests fails because it involves an example with two *differem* experiments. He does admit that it is somewhat doubtful 'whether we can apply both tests to the same sample' because if we do we are raising the level of k, and may raise it above the level we are willing to accept. But this remark misses the essential point: If one researcher is entitled to regard an hypothesis as falsified under given data and another (or the same) researcher is entitled to regard it as corroborated by the data under the very same rule, there is something quite wrong with this rule. It has as much value as a deductive rule that sanctions contradictory inferences from the same premisses. If Neyman's example constituted a rather special, isolated anomaly for Gillies's rule, we might tolerate it in the hope that minor modifications of the rule can resolve it. But as we have seen, it is an instance of a pervasive difficulty, and Gillies's position seems to be untenable. In fact, rather strong arguments against the view that both types of tests are valid in general can be found in Gillies's argument for the soundness of his rule. Suppose that the hypothesis H mentioned in the second paragraph states that Z has a unimodal and symmetrical distribution. The falsification class C'' involves a small interval centred about the mode. Such classes are flatly declared by Gillies on *p.* 241 to be 'counter-intuitive', and he remarks that the purpose of introducing the notion of a falsifiable distribution is to eliminate them. But his rule invites them, as shown above. In his discussion of the uniform distribution on *pp.* 242 and 243 Gillies argues, in effect, that no falsification class for the hypothesis that W is uniformly distributed of the type corresponding to C'' is valid. He further suggests that the test corresponding to C' is not valid as well. It should be obvious even to Gillies that both sorts of tests are not valid. And there is no hope of modifying the rule to rule out classes like C'' in favour of those like C', because the latter clearly violate the spirit of the rule when the distribution of Z is U-shaped.

Another serious defect of Gillies's rule—on his own criterion of adequacy—

is that it conflicts with established procedure in problems involving goodness of fit tests. For such problems a statistic D is defined over the outcomes of the experiment such that the statistic provides a linear ordering of the degree to which the observed frequencies diverge from the frequencies that would be expected if the hypothesis H is true. More precisely, if O and O' are possible outcomes of the experiment, $D(O) > D(O')$ if and only if O' supports H better than O. If the experiment is large and H true, D is approximately Chi-Square distributed. H is rejected if and only if the discrepancy between H and the frequency distribution generated from the actual result O^* of the experiment is so large that $P(D \geqslant D(O^*)) < \alpha$, where α is a 'small' probability. The usual value of α is .05, although a value of .01 or less is sometimes used. But Gillies's rule would have us reject H not only for high values of D, but also for very small values of D. In other words, it recommends a two-tailed test which would have us regard H as falsified if the data provides a very close fit to the theoretical distribution hypothesised by H. But virtually all standard textbooks treat this test as a one-tailed test in the manner sketched above. Gillies appears to have only one influential ally, namely, R. A. Fisher. As Gillies pointed out in correspondence with me, Fisher clearly states in his [1958], *p*. 81, that if O^* is so close to what we would expect if H is true that $P(D \geqslant D(O^*)) \geqslant .999$, H is 'as definitely disproved as if P had been .001'. It is not clear that Fisher would endorse the sort of two-tailed test Gillies's rule recommends, but there are indications that he would. This is not to say that Fisher can be cited as an authority for his rule, for Fisher's treatment of the Lady Tasting Tea Experiment in his [1966] involves the use of a one-tailed test for a symmetric unimodal distribution, *i.e.* precisely the sort of test Gillies's rule would condemn. It would be pointless for me to trot out a list of journeyman statisticians who disagree with Fisher on this point, for it includes virtually all authors of reputable statistics texts. The general consensus among those who do discuss the occurrence of very high values of $P(D \geqslant D(O^*))$ is that such values do not count against the hypothesis under test but rather tend to discredit the data or the assumptions underlying the use of the Chi-Square criterion. For example, in his discussion of Chi-Square goodness of fit tests Tippett writes

> the appropriate critical region . . . for the ordinary test of significance is that containing values of χ^2 greater than that at which an ordinate cuts off a single tail of .05 or whatever other level is chosen. A value of P very near to 1.0 (say, greater than .99) corresponds to *no significant difference between experiment and hypothesis*, but it indicates an agreement that is too close for an experience subject to random variations, and it usually results from some error in the application of the theory. ([1952], *p*. 129, my italics.)

In the language of significance testing the phrase 'no significant difference . . .' means that the data do not count against the hypothesis. A concrete illustration of this issue is provided by the controversy sparked by Fisher's paper 'Has Mendel's Work Been Rediscovered?' ([1936]), in which he pointed out that Mendel's experimental data as a whole fits the theoretical model so closely that the probability of a worse fit is .99993. Fisher suggests that if independent corroboration of Mendel's theory were not available, Mendel's data would render it highly suspect. The eminent geneticist A. H. Sturtevant criticised

Fisher's analysis on the ground that 'it is possible to question the logic of the argument that a fit is too close to expectation', since it 'comes perilously close to saying that no such experiment can be carried out . . .'.[1] Sturtevant then offers possible explanations of the extreme closeness of fit which involve some ways in which Mendel's data might have become biased. Another biologist who has considered this issue writes, in a highly respected text on genetics,

> Mendel's results have been taken as an example of data which give an exceptionally high probability with the χ^2 test. In general, P values within the range .1 to .9 cause no suspicion and indicate that the observations fit the hypothesis under test, although they may of course fit other hypotheses. Values of .05 or less are conventionally said to indicate a real discrepancy . . . Conversely, values of .95 or more indicate an abnormally close fit with hypothesis, while probabilities as high as .98 or .99 would certainly arouse suspicion that either consciously or unconsciously, the data had become biased in favour of the hypothesis.[2]

It must be noted that there may be some tests which use a Chi-Square criterion, and are similar to goodness of fit tests, for which a high P value would count against the hypothesis tested. But such tests would differ from goodness of fit tests in that they involve a precise alternative hypothesis which predicts, in effect, that an abnormally close fit to the hypothesis under test is to be expected.

A related criticism of Gillies's paper is that he does not consider the modified Fisherian theory of testing which is used by the vast majority of researchers in the social and bio-medical sciences who deal with statistical testing. This theory is an amalgam of Fisher's theory of significance and the Neyman-Pearson theory, but its logic is broadly Fisherian in character. On this theory alternative hypotheses are considered when they are available, but they function only to determine the statistic to be used and whether or not the test is one- or two-tailed. In other words, considerations of power do not enter into construction of the rejection criterion; only P values do. The theory countenances tests where no precise alternative hypothesis is available, unlike the Neyman-Pearson theory. I say its logic is broadly Fisherian in character because Fisher never formulated a general, consistent and systematic rationale for the theory of significance testing, so that researchers could only try to treat problems in the manner suggested by Fisher's concrete examples (without always agreeing with him, as shown above). A brief sketch of this modified Fisherian theory is provided by Cramer.[3] To test a statistical or probabilistic hypothesis H an experiment E has to be devised such that a statistic D which provides a linear ordering of the divergence of each possible outcome of E from what would be expected if H is true (and the alternative false if there is a precise alternative) is definable over E. Moreover, it must be possible to compute the distribution function of D on the assumption that H is true. If O and O' are possible outcomes of E, $D(O) > D(O')$ if and only if O' supports H better than O does. The basic idea of the theory is that if $P(D \geqslant D(O))$ is 'small' on the assumption that H is true, then O counts against H to such a degree that we are entitled to reject H if O is the observed outcome of E. Note that on this theory, as opposed to Gillies's,

---

[1] Sturtevant [1965], p. 13.

[2] Whitehouse [1967].

[3] Cramer [1946] p. 525.

H is not rejected on the grounds that the hypothesis that *D* has the distribution *H* implies that it has is 'falsified'. *D* is used as a criterion to directly test *H*. Thus this theory is free of the difficulties faced by Gillies's rule. It is unscathed by the arguments of Neyman discussed by Gillies in section 4 of his paper, because all sorts of judgments of evidential support are incorporated in the definition of a *D*-function for a particular problem. It does not generate any of the counter-intuitive falsification classes mentioned by Gillies, although there may be several candidates for *D*-functions which do not strictly agree.

The modified Fisherian theory of significance testing is not free of difficulties, however. It is mistaken in its view that $P(D \geqslant D(O))$ given *H* is an index of the evidential weight of data *O*, as I have shown in an unpublished paper on the logic of significance testing. My own view on the general problem of providing a falsifying rule for probability statements is that no solution exists if subjective appraisals of the credibility of hypotheses in relation to all one knows and judgments of inductive probability are ruled inadmissible ingredients of a solution. If (which I doubt) Popper's claim, cited by Gillies on *p*. 232, that 'the physicist knows well enough when to regard a probability assumption as falsified' does hold water, the rule that accurately describes the behaviour of physicists would incorporate the subjective ingredients mentioned above. More precisely, it would conform to a Bayesian analysis.

STEPHEN SPIELMAN
Lehman College of the City University of New York

REFERENCES

Cramer, H. [1946]: *Mathematical Methods of Statistics.*
De Maré, J. [1972]: 'A Comment on Gillies' Falsifying Rule for probability Statements', *British Journal for the Philosophy of Science*, **23**, p. 335.
Fisher, R. A. [1936]: 'Has Mendel's Work Been Rediscovered?', *Annals of Science*, **1**, pp. 115–37.
Fisher, R. A. [1958]: *Statistical Methods for Research Workers.*
Fisher, R. A. [1966]: *Design of Experiments.*
Gillies, D. A. [1971]: 'A Falsifying Rule for Probability Statements', *British Journal for the Philosophy of Science*, **22**, p. 231–61.
Sturtevant, A. H. [1965]: *A History of Genetics.*
Tippett, L. H. C. [1952]: *The Methods of Statistics.*
Whitehouse, H. L. K. [1967]: *The Mechanism of Heredity.*

ON NEYMAN'S PARADOX AND THE THEORY OF STATISTICAL TESTS

In a recent [1971] paper D. A. Gillies has introduced a falsifying rule for probability statements and has attacked the Neyman–Pearson theory of statistical tests employing the use of alternative hypotheses, from a number of points of view. This work has already attracted critical comment.[1] The main purpose of the present paper is to consider once more Neyman's objections to the type of argument advanced by Gillies, as expounded in his [1952].[2] In particular Neyman produced an example of two statistics z and ζ related in such a way that if used to test some hypothesis H, the following possibilities would obtain: (*i*) both the z-test and the ζ-test might corroborate H, (*ii*) if the z-test refutes H the ζ-test must necessarily corroborate H, (*iii*) if the ζ-test refutes H the z-test must necessarily corroborate H. In no case will both tests lead to refutation. These paradoxical results lead Neyman to the view that some additional criterion must be introduced to select one statistic as against another as a suitable basis for testing a statistical hypothesis. Such a criterion is afforded by the use of alternative hypotheses in the standard Neyman–Pearson theory, which effectively selects that test which gives the largest probability of rejecting the hypothesis when it is in fact false, *i.e.* an alternative hypothesis is true.

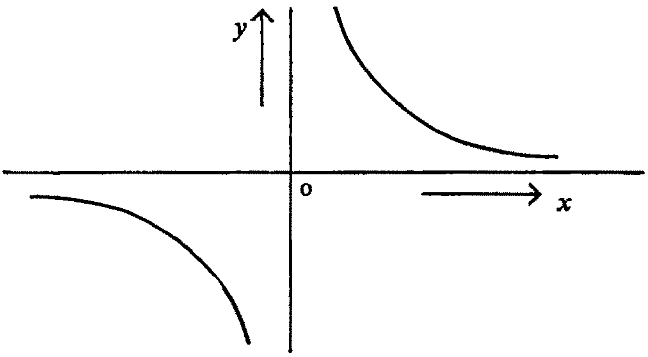
---

[1] See De Maré [1972], Gillies [1972] and Spielman [1974].

[2] Neyman [1952], pp. 43–54.

In Neyman's own version of his paradox the statistics z and ζ were not functionally related, but were connected by the inequality $|z-\zeta| < 1$. We begin by noticing that Neyman's paradox can be formulated much more simply and much more generally by considering the properties of two 1-dimensional random variables each of which is a function of the other.

Let x be a random variable defined on the real line, and we suppose that x has a continuous unimodal 'bell-shaped' distribution which is falsifiable according to Gillies's rule, so that the range of x can be divided into a 'head' and two 'tails', the two tails constituting the falsification class associated with the distribution. We assume for simplicity that the median of the distribution is located at $x = 0$, but do not assume that the distribution is necessarily symmetric about the median. We denote the distribution function of x by $F(x)$. Now introduce a new random variable $y = \phi(x)$, where the graph of ϕ is of the form illustrated in the accompanying diagram.



y decreases monotonically with x in the ranges $-\infty < x < 0$, $0 < x < \infty$ and tends to zero as x tends to $\pm\infty$, but suffers an infinite discontinuity at $x = 0$, so that the range of y is the real line minus zero.

Denoting the distribution function of y by $G(y)$, it is easy to see that

$$G(y) = \begin{cases} 3/2 - F(x), & x > 0 \\ \tfrac{1}{2} - F(x), & x < 0 \end{cases} \tag{1}$$

The corresponding frequency functions $g(y)$ and $f(x)$ are clearly related by

$$\begin{aligned} g(y) &= -f(x)\frac{dx}{dy} \\ &= -f(\phi^{-1}(y))\frac{d}{dy}\phi^{-1}(y) \end{aligned} \tag{2}$$

Now suppose that $\phi(x)$ is chosen so that x and y have identical distribution functions. The required invariant transformation can always be found and is simply determined by

$$y = \phi(x) = \begin{cases} F^{-1}[3/2 - F(x)], & x>0 \\ F^{-1}[\tfrac{1}{2} - F(x)], & x<0 \end{cases} \tag{3}$$

But the paradox is now apparent since the falsification class for y is clearly located in the acceptance region for x, and vice versa, when Gillies's rule is applied to the identical distributions for x and y.

As a very simple and indeed degenerate example consider the frequency function given by

$$f(x) = \frac{1}{4 \ln L}\frac{1}{|x|} \tag{4}$$

defined over $-L < x < -\frac{1}{L}, \frac{1}{L} < x < L$, where L is a large positive quantity introduced to secure convergence. If we calculate the corresponding distribution function and substitute in (3) we obtain the invariant transformation $y = 1/x$.

The same transformation is actually an invariant transformation for the Cauchy distribution with frequency function

$$f(x) = (\pi(1 + x^2))^{-1} \tag{5}$$

as may readily be verified using equation (2). This is the example treated already in Spielman [1974]. Our analysis confirms the point that Spielman makes, namely that Neyman type paradoxes can be generated at will by appropriate transformations of random variables.

However Gillies rejects the logic of Neyman's argument by considering an example, drawn from deterministic physics, of two tests of Galileo's law for the acceleration of freely falling bodies *in vacuo*, which he claims are related in the same way as Neyman's tests, but which are equally valid as tests and do not necessitate choosing one as more valid than the other. I wish to point out a flaw in Gillies's analogy which seems to me to vitiate his conclusion, namely that in the statistical case also there is no logical necessity to choose between tests related in the manner of Neyman's example.

Gillies's argument can be stated quite simply as follows (the involved experimental details described in his paper serve somewhat to obscure the principle involved). Let a physical hypothesis H state that some observable ξ has the value ξ_0. Consider two tests T_1 and T_2 such that T_1 compares ξ with $\xi_0 - \epsilon$, where ϵ is some prescribed positive quantity, and T_2 compares ξ with $\xi_0 + \epsilon$. Clearly if T_1 shows that $\xi < \xi_0 - \epsilon$, so that H is refuted, then T_2 must show that $\xi < \xi_0 + \epsilon$, so that H is corroborated, while if T_2 shows that $\xi > \xi_0 + \epsilon$, so that H is refuted, T_1 must show that $\xi > \xi_0 - \epsilon$, so that H is corroborated. We can of course have $\xi > \xi_0 - \epsilon$ and $\xi < \xi_0 + \epsilon$, so that both T_1 and T_2 may lead to corroboration, but it is impossible to have simultaneously $\xi < \xi_0 - \epsilon$ and $\xi > \xi_0 + \epsilon$, so we cannot have both tests leading to refutation.

So far the situation is just the same as in Neyman's example of the z-test and the ζ-test. But now we notice an important difference. In the T_1 and T_2 tests considered above, if one test corroborates H it is not inconsistent with H being false, but if one test refutes H it is definitely inconsistent with H being true. Hence in the cases where T_1 and T_2 conflict the choice between them is quite unequivocal, we must always follow the test which leads to refutation. There is no need here for any further criterion to help us decide.

However, in the statistical example the asymmetry noticed for T_1 and T_2 does not apply. If either the z-test or the ζ-test corroborates the statistical hypothesis H, H may still be false, but if either test refutes H, there is no inconsistency in H being, in fact, true. Thus we are no longer compelled to accept the test leading to refutation as in the case of T_1 and T_2. Furthermore we cannot reasonably resolve the situation by adopting as a methodological rule the acceptance of the test leading to refutation (referred to by Neyman as Principle II), since this would be an arbitrary procedure which would actually conflict with the recommendations of both Neyman and Gillies, who agree, for different reasons, that the z-test should be accepted, although Principle II would tell us to reject it if the ζ-test led to refutation. (Incidentally I regard Gillies's criticism of Neyman's own attack on Principle II as entirely valid.) It is sometimes assumed that Principle II reflects a Popperian methodology, but this is quite wrong since Popper demands in science refutability, not refutation. The question of whether a particular test is 'accurate' enough to achieve the refutation of a hypothesis is left quite open by Popper as a matter for separate discussion and evaluation with respect to the conditions of the test.

In order then to choose between the z-test and the ζ-test in cases where they conflict, some additional criterion must be introduced, and Neyman's contention that this should be rationally based on the consideration of alternative hypotheses is certainly a reasonable one, even if practical difficulties, stressed by Gillies, make the implementation of the underlying philosophy a somewhat arbitrary procedure.

It may be suggested that the paradoxical conflict between two statistics envisaged by Neyman would never arise in a practical statistical investigation in which only one statistic would normally be employed, but such a pragmatic argument misses the thrust of Neyman's reasoning, which was to stress in principle the inadequacy of assessing tests using properties of the distribution function for the statistic on which the test is based, specified by the hypothesis being tested. It is on the matter of principle that Gillies's objection to Neyman's argument breaks down, by employing an incorrect analogy with situations arising in deterministic physics.

Spielman also criticises Gillies's objection to Neyman on the ground that Gillies's deterministic example is one with two *different* experiments, whereas Neyman's statistical example involves only a single experiment. However, we can always consider two experiments as a single combined experiment. Thus

Spielman's counter-argument is not in itself very convincing. I believe that the discussion above goes rather more deeply into the issues involved here.

An approach to the theory of statistical testing which differs from both Gillies and Neyman–Pearson is the Bayesian one. It is interesting to notice how a Bayesian analysis avoids the pitfall of the Neyman paradox.

Consider a range of hypotheses determined by a continuous parameter α, where $\alpha = 0$ corresponds to the hypothesis H being tested. To make the mathematics simple we choose H as specifying $-\epsilon < \alpha < \epsilon$ and take $\epsilon \rightarrow 0$.

Denote by P_T the probability that H is true and suppose the random variable x, whose distribution is specified by H, is observed to have a value in the range x to $x + dx$. Then we can readily deduce from Bayes's theorem

$$P_T = \frac{1}{1 + \dfrac{n(x)}{m(x)}} \tag{6}$$

$$\text{where } m(x) = \int_{-\epsilon}^{\epsilon} P\,(x/\alpha)\, P\,(\alpha)\, d\alpha \tag{7}$$

$$\text{and} \quad n(x) = \left(\int_{-\infty}^{-\epsilon} + \int_{\epsilon}^{\infty}\right) P(x/\alpha)\, P(\alpha)\, d\alpha \tag{8}$$

$P(x/\alpha)$ and $P(\alpha)$ are the conditional frequency function for x and the prior frequency function for α respectively.

We write $\quad P(\alpha) = \tfrac{1}{2}(\delta(\alpha) + \chi(\alpha)\,) \quad (9)$

where $\chi(\alpha)$ is regular at $\alpha = 0$, $\delta \equiv \delta$-function and $\int_{-\infty}^{\infty} \chi(\alpha)\, d\alpha = 1$.

Thus before measuring x, $P_T = \tfrac{1}{2}$, so our prior assumption is that there is equal probability of H being true or false.

Substituting from (9) in (7) and (8) gives

$$m(\chi) = \tfrac{1}{2}\, P(x/0) \tag{10}$$

$$n(x) = \tfrac{1}{2} \int_{-\infty}^{\infty} P(x/\alpha)\, \chi\,(\alpha)\, d\alpha \tag{11}$$

Of course, to achieve a test of H, using the statistic x, the ratio $n(x)/m(x)$ must vary from much less than unity in the acceptance region to much greater than unity in the critical region. Since in the Bayesian analysis P_T is determined by the ratio of two frequency functions, a mere change in random variable cannot alter the value of P_T, so no paradox (conflicting values of P_T) can arise. Thus introducing a change of variable $y = \phi(x)$ of the form previously discussed, we find at once (*cf.* equation (2))

$$\left.\begin{aligned} m(x) &= -\,m'(y) \,.\, \frac{dy}{dx} \\[2ex] n(x) &= -\,n'(y) \,.\, \frac{dy}{dx} \end{aligned}\right\} \tag{12}$$

where $m'(y)$ and $n'(y)$ are given by (10) and (11) with $P(x/\alpha)$ replaced by $P'(y/\alpha)$, the conditional frequency function for the variable y. Now suppose $y = \phi(x)$ is an invariant transformation for $P(x/0)$ so that $m'(y) = m(y)$.

Then from (6) and (12)

$$P_T = \frac{1}{1+\dfrac{n'(y)}{m(y)}} \tag{13}$$

This is the situation which led previously to the Neyman paradox, when we based our criterion for rejecting H purely on properties of the likelihood function $m(x)$, for which corresponding values of x and y as argument give different values.

But equations (6) and (13) imply that corresponding values of x and y will always give the same value for P_T. The paradox is thus avoided in the Bayesian analysis by allowing for the non form-invariance of $n(x)$, which is the frequency function for observing x on *some* alternative hypothesis. Incidentally the connection with the Neyman–Pearson theory is apparent where the rejection of the paradoxical statistic is also associated essentially with properties of the function $n(x)$, which is clearly related to the power function of the test.

In conclusion I add a few general remarks on the source of the difficulties associated with Gillies's rule.

These all stem from the definition of the relative likelihood function $l(x)$ for the case of a continuous distribution given on page 234 of Gillies's paper.

He writes $$l(x) = f(x)/f_{max} \tag{14}$$

and formulates his rule in terms of properties of $l(x)$, *e.g.*

$$l(x) < l_0 \tag{15}$$

for x in the falsification class C.

Now (15) can be rewritten as

$$f(x_1)\,(\delta x)_{x=x_1} < l_0\, f(x_2)\,(\delta x)_{x=x_2} \tag{16}$$

where $f(x_2) = f_{max}$ and $x_1 \in C$

and $$(\delta x)_{x=x_1} = (\delta x)_{x=x_2} \tag{17}$$

(16) compares the probabilities of two sets of events defined to lie in positive ranges $(\delta x)_{x=x_1}$ and $(\delta x)_{x=x_2}$ about the values x_1 and x_2 for the random variable x. Equation (17) says that these two ranges are numerically equal. But equation (17) is a *completely arbitrary assumption* which Gillies makes no attempt to justify, and which leads directly to the Neyman paradox since it is clearly not invariant under a change of random variable.

Clearly what is required is to introduce an invariant[1] measure μ on the real line in terms of which (17) can be replaced by

$$\mu\,[(\delta x)_{x=x_1}] = \mu\,[(\delta x)_{x=x_2}] \tag{18}$$

This is most conveniently expressed in terms of a canonical random variable ξ for which $\mu\,(\delta\xi) = |\,\delta\xi\,|$ so that

$$\mu\,(\delta x) = \left|\frac{d\xi}{dx}\right|\,\delta x \tag{19}$$

---

[1] By invariant I mean that the same set of events as specified by different random variables have the same measure.

and equation (18) reads just

$$\left|\frac{d\xi}{dx}\right|_{x=x_1}(\delta x)_{x=x_1} = \left|\frac{d\xi}{dx}\right|_{x=x_2}(\delta x)_{x=x_2} \tag{20}$$

This equation is invariant under change of random variable, so no Neyman type paradoxes arise.

Thus Gillies's rule must be replaced by the concept of a distribution which is falsifiable with respect to a canonical random variable. But at this point the question of how we select the canonical random variable will arise and this will lead us to introduce qualitative considerations for using a test statistic which do not depend solely on the distribution function for the statistic. This leaves open the question of whether these additional considerations should involve introducing alternative hypotheses, but at all events Gillies's original rule cannot escape the Neyman paradox. The best that may be said, following Gillies, is that the statistic used should measure quantities of practical interest, or be closely related to such quantities. This might still allow some flexibility in re-scaling,[1] but would eliminate transformations which alter the linear ordering of the statistic, and it is just such transformations that lead to Neyman's paradox. With such an addition Gillies's rule may provide a viable alternative to other approaches to the difficult problem of testing statistical hypotheses.

M. L. G. REDHEAD
Chelsea College,
University of London

REFERENCES

CAMPBELL, N. R. [1957]: Foundations of Science.

DE MARÉ, J. [1972]: 'A comment on Gillies's Falsifying Rule for Probability Statements', *British Journal for the Philosophy of Science*, **23**, p. 335.

GILLIES, D. A. [1971]: 'A Falsifying Rule for Probability Statements', *British Journal for the Philosophy of Science*, **22**, pp. 231–61.

GILLIES, D. A. [1972]: 'Reply to De Maré', *British Journal for the Philosophy of Science*, **23**, pp. 335–36.

NEYMAN, J. [1952]: *Lectures and Conferences on Mathematical Statistics and Probability.*

SPIELMAN, S. [1974]: 'On the Infirmities of Gillies's Rule', *British Journal for the Philosophy of Science*, **25**, pp. 261–265.

---

[1] For a general discussion of re-scaling in the theory of measurement the work of N. R. Campbell ([1957], part II) may be usefully consulted.

SITUATIONAL DETERMINISM IN ECONOMICS

This article was written in response to one published under the same title by Spiro J. Latsis in this *Journal*.[1] The title puzzled me when I first saw it, but Latsis explained it in the first sentence, where he contrasted 'economic determinism' with 'economic behaviouralism' and referred to them as 'two rival research programmes in a major branch of modern microeconomic theory, in the theory of the firm'. Since I had, a few years earlier, produced a piece on

---

[1] Latsis [1972].

'Theories of the Firm: Marginalist, Behavioural, Managerial',[1] I realised quickly that Latsis and I were either rivals or allies in the same controversy. In the end I concluded that he was on my side more often than not. Still, on some critical questions we seem to disagree and it is, I believe, worth looking for the reasons. Not wanting to impose on the reader a reading or re-reading of our earlier pieces, I shall try to make the present article intelligible to those who are tuning in only now. A conscientious reader, however, will have to go back to Latsis's article to check my interpretation of his position.

1 *Nomenclature*

Economic determinism is Latsis's term for the general principles that have characterised the neoclassical programme in the theory of the firm, which attempts to explain market phenomena as the results of actions of rational agents who, in an attempt to maximise their profits, respond to the 'logic of the situation'. Situational determinism seems to be the same as marginalism inasmuch as this term refers to the logical procedure of firms equating marginal revenue and marginal cost in attempting to maximise profits.

Latsis's term, or rather a part of it, is derived from a forceful dictum by Popper: 'The analysis of the situation, the situational logic, plays a very important part in social life as well as in the social sciences. It is, in fact, the method of economic analysis.'[2] The main point regarding this simple theory of the firm is that it is based on the postulate of a single motive or objective of a single decision-maker, excluding therefore all the complications that can arise from the presence of several conflicting goals motivating several (more or less cooperating) decision-makers within a bureaucratic organisation.

The behavioural theory of the firm substitutes for the single goal of profit maximisation a collection of different goals pursued by a collection of people in executive positions in a business organisation. (Latsis proposes that 'it may be advisable to develop behaviouralist ideas first in their application to a single decision maker and later to a complex decision making structure'; he thus opposes the 'combination of behaviouralism and organisationalism'.[3] But the main tenet of behaviouralism is that the objectives of the decision-makers (and the techniques of resolving conflicts among them) should not be postulated but should first be ascertained by empirical research in actual firms, particularly by observing decision-making processes in reality. That the observations should be aided by working hypotheses and designed to develop a theory 'with generality beyond the specific firms studied'[4] justifies the term 'behavioural' as distinguished from 'behaviourist'.

Latsis mentions but does not give major billing to the third approach, the managerial theory of the firm. While the managerial theory hypothesises a few other specified goals[5] in addition to the goal of profit maximisation, it postulates that there are rational ways to reconcile the chosen maximanda with one another, for example, by making the additional managerial goals subject to the constraint of some minimum benefits to the owners of the business. In one way or another, the accommodation of stated objectives besides that of maximum money profits in a model of the firm may justify placing managerial theories together with marginalist theories or, in Latsis's terminology, with 'situational determinism'. In any event, I shall for the present discussion accept Latsis's suggestion of only two rival research programmes in the theory of the firm.

---

[1] Machlup [1967]. I shall cite an immodestly large number of my own writings in order to facilitate the work of those who want to go back and look into the development of the ideas behind this discussion.

[2] Popper [1945], *volume* 2, p. 97.

[3] Latsis [1972], footnote 3 to p. 229.

[4] Cyert and March [1963], p. 2.

[5] Baumol has proposed maximisation of sales or of the growth rate of sales as likely managerial objectives (Baumol [1959], p. 49, and [1962], p. 1086). Oliver Williamson added maximisation of expenditures for staff personnel and for managerial emoluments (Williamson [1964], pp. 38–60).

2 *Rivals in What Sense?*

Latsis, in regarding marginalism and behaviouralism as rival theories, or rival research programmes, accepts the claims of the behaviouralists who reject the marginalist programme (or paradigm) and offer their own as a better or more promising substitute. I submit, however, that the two programmes are rivals only as long as they are supposed to serve the same purpose.

For example, if both the marginalist and the behaviouralist programmes are understood as systems designed to explain and predict the actions of particular business firms, one may regard them as rivals and may judge their comparative performance in this task. It is no doubt true that some writers in the marginalist camp have presented their theory as an instrument performing that task. However, this is not the main purpose of the marginalist theory of the firm and, if it is used for this purpose, it will be rather unreliable and will utterly fail in far too many cases.

The proper use of marginalist theory is to explain and predict the direction of changes in prices and outputs of particular products consequent upon particular events or changes in conditions. In the model that establishes a connection between these events (or changes in conditions) and the resulting changes in prices and outputs the firm is merely a link (bridge, transmission belt, transformer, automaton, or robot) between causes and effects. For this function a behavioural theory of the firm is not suitable, because it implies that the behaviour may be too varied, too idiosyncratic, too capricious (yes, too human) to be a reliable link in the chain from typical causes to typical effects. As Latsis correctly states, 'the burden of explanation shifts to the behavioural characteristics of the acting agent' (p. 229). Yet since the spectrum of such characteristics includes too many nuances and we would not know which one might apply in any concrete case, we could not hope to get much help for the stated purpose.

To be sure, the model employing this simple link-element called 'firm'—simple as long as it works merely as a robot with a one-track mind (or calculator)—can be used safely only under certain conditions ('used', that is, for purposes of explaining and predicting in particular historical situations). The most important of these conditions is that the number of real-world firms is large and that none of them has a dominant position. However, even when the model cannot be used safely (that is, with confidence that it will yield correct predictions), it will usually not work any worse than the behavioural programme with its choice of, say, 57 standard models and innumerable variants. Indeed, since the behavioural programme will rarely ever be sufficiently specified to tell us exactly which of the many possible behaviour models should be used in a particular case (particular as to industry, firm, product, place and time) we shall be at a loss trying to make the right selection. To 'misapply' the simple neoclassical model in situations in which it is not supposed to give the right answers might still give us less inaccurate results—and would surely be much less cumbersome.

Latsis understands this full well. This is what he has to say: 'In general there will be an indefinite variety of alternative decision processes for the same situational and motivational picture and each may lead to very different decisions being adopted. The actual decision may depend on more or less idiosyncratic rules or on rather stable rules related to various personality traits of the decision makers or to their social background or to other social-psychological forces' (p. 228). I am afraid, one cannot be certain that observation of the decision processes will result in any generalisations at all. This has been recognised by a student of the problem of 'simulated' behaviour, though he has not given up the hope that we shall end up with a manageable number of 'basic' models: 'In order to obtain adequate descriptions of the behaviour of individual firms, *it might conceivably be necessary to have as many models as there are firms*. . . . We doubt that this will be the case, however. Probably a relatively small number of basic models . . . will adequately describe the relevant aspects of each individual firm's actions.[1]

The loss of generality is recognised by all behaviouralist theorists. Cyert and March, the authors of the authoritative text of the school, admit this in the following statement: 'The behaviour theory of the firm relaxes considerably the requirements of simplicity in the models, expands considerably the phenomena that can be considered explicitly, and reduces considerably the need for aggregation of phenomena into summary variables. As a result, the theory gives up some attributes of generality.'[2]

I have stated that marginalist theory will strictly apply only when competition is vigorous and therefore effective (Machlup [1967], p. 15). Latsis believes that this tells us merely 'that the perfectly competitive model is only applicable where it can be applied' (p. 212). I wonder whether this does full justice to my statement:

> The simple marginal formula based on profit maximization is suitable where (*1*) *large groups* of firms are involved and nothing has to be predicted about particular firms, (*2*) the effects of a *specified change* in conditions upon prices, inputs, and outputs are to be explained or predicted rather than the values of these magnitudes before or after the change, and nothing has to be said about the 'total situation' or general developments, and (*3*) only *qualitative answers*, that is, answers about directions of change, are sought rather than precise numerical results (1967, p. 31).

One more illustration of the use of the simple, general, marginalist theory may be helpful.

> How will the price of wheat be affected by a report of a drought; by a reduction in the import quota for wheat; by a reduction in the rate of interest; by an increase in freight rates; by an announcement that the ice cover on the Great Lakes will delay the opening of shipping for several weeks? These questions can be answered with the aid of general propositions of economic theory. The answers do not presuppose that the economist

---

[1] Cohen [1960], pp. 536–7 (my italics).

[2] Cyert and March [1963], p. 281.

> knows any wheat dealer personally, let alone his psychological make-up. They do presuppose, however, that the economist has constructed an ideal type of dealer conduct. Its main feature is that dealers would rather make more money than less. This imputation of the profit motive to anonymous characters—'intervening variables' between, say, a newspaper report and a quotation of a higher price on the wheat exchange—is necessary for a full understanding of the causal connections.[1]

A research programme designed to result in theories that explain and predict the actions of particular firms can never compete with the simplicity and generality of the marginalist theory which, being based on the construct of a fictitious profit-maximiser, cannot have the ambition to explain the behaviour of actual firms in the real world. The two research programmes are designed for different purposes and are therefore not real rivals.

3 *The Compromise Regarding Oligopoly*

Latsis believes that an 'appraisal of the rival research programmes is premature' (p. 233). I am quite willing to wait and give the behaviouralists more time to come up with a set of models and a set of directives and specifications telling which model to apply to which kind of firm and which kind of situation. When they do and when their models give us results superior to the 'naive' marginalist model, we shall be able to judge whether the improvement (greater accuracy, numerically determinate results) is worth the cost (need for much more information).

On one point Latsis thinks he is ready for a judgment: 'But I can say now that I see no prospect for a compromise solution whereby situational determinism retains most of the theory of the firm but concedes oligopoly to behaviouralism' (p. 233). His reasons are that each of the two approaches entails a 'pervasive methodological view', and that behaviouralism, if it has a special function in markets where sellers are few, may show also that 'even under conditions of apparently severe competition there are systematic patterns of divergence from optimising behaviour' (pp. 233–4). These are no cogent reasons. It would have to be shown that the 'divergence from optimising behaviour' in situations of effective competition is large enough to warrant the sacrifice of simplicity imposed by the behavioural programme. If a simple technique can yield good results in a large range of cases while a much more complicated and costlier technique is needed for some, should this be a compelling reason for using the costlier technique throughout? I think not. If consistency in approach is expensive or even extravagant, it is no longer a virtue.

The 'compromise solution' mentioned by Latsis preceded by several decades the proposals for a behavioural theory of the firm. One of the early proponents of the 'compromise' was Vilfredo Pareto. In his view 'pure economics' cannot yield a unique solution in the case of few sellers and 'observation' will show a variety of possibilities.[2] I cited Pareto when I discussed 'Oligopolistic Indeterminacy' in a book published in 1952, and went on to cite Alfred Schutz on the difference in analysing the effects of the behaviour of anonymous masses on the

[1] Machlup [1969], pp. 301–302.

[2] Pareto [1927], pp. 601–2.

one hand, and the conduct of particular individuals on the other,[1] I concluded that for an analysis of the effects of mass behaviour we can often do with far less information than we may need for an analysis of the actions of a few particular persons. (Where the outcome depends on judgments and decisions by a few individuals the selection of the appropriate model would require characterological research, or at least investigation of the 'voting records' of the individuals in question and of their influence in the hierarchy of decision-makers in the organisation.) It would be a strange methodological principle indeed that were to command us to seek the heaps of information that may be required in some cases also for problems where they are not needed.

Incidentally, to admit that the marginalist approach may fail to yield solutions in cases where the number of decision-makers is relatively small is not to concede that the behavioural approach has succeeded or is likely to succeed any better. It has not yet furnished unique solutions and accurate predictions. I have my doubts that it ever will, but I keep an open mind.

4 *Three Crucial Confusions*

In his critique of what he calls situational determinism, Latsis does not guard himself and does not warn his readers sufficiently against some of the traps in which most critics of neoclassical economics have been caught. Three confusions are rather common in the literature:

1. The confusion between action and reaction. To explain or predict anybody's actions is a far bigger job than to explain or predict his *reactions* to certain types of events or changes of conditions. Indeed, the number of possible actions is so large that the task of predicting them is almost impossible. In my earlier writings I have stressed the difference between entrepreneurial, innovative actions and typical responses to typical stimuli.[2]

2. The confusion between reactions and effects of reactions. It would be practically impossible to ascertain the reactions of millions of people to specified types of events; even to check large samples of supposed reactors would be difficult where neither the stimulus nor the response is physical. Where the stimulus is only a message, a piece of information, a becoming aware of something, and where the response is a decision (usually embodied again in some sort of message or instruction) 'direct' observation of the decision and its connection with the stimulus may be virtually impossible. To ascertain the *effects* of the decision is something quite different.[3] One can observe the effects of business decisions by consulting the records of certain changes in market prices or production volumes.

3. The confusion between the effects of reactions of particular persons and the effects of mass reactions or, more correctly, of assumed typical reactions of imaginary persons. The essential difference here is between actual people (living persons or historical personalities) and idealised, *fictitious* people invented (constructed) for the purpose of deducing from their assumed (built-in) reactions effects of the type observed in reality.[4]

---

[1] Machlup [1952], pp. 417, 419.

[2] Machlup [1967], pp. 7, 8, 13, 27.

[3] Machlup [1967], pp. 7, 8; [1970], pp. 123, 139.

[4] Machlup [1963], pp. 131–4, 142; [1967], pp. 6–8, 27; and [1970], p. 139.

Compounding all three pairs of contrasts into one, we confront the huge difference between (*a*) observable actions of real people and (*b*) typical effects of imagined (postulated) reactions of imagined (postulated) people to typical events or changes in conditions. Both the typical effects and the typical events to which they are attributed correspond to classes of observable changes in recorded data.

I cannot claim that Latsis himself is a victim of any of these confusions, but he does little to make his readers beware of the snares. Regarding the confusion between *action* and *reaction*, I note the almost complete absence of the words reaction and response in a discussion in which words such as actions, courses of action, decisions, decision procedures, choices, and behaviour occur several times on every page. Only once does Latsis come to grips with the problem: when he speaks of 'changes in variables in response to variations in parameters' (p. 213). Yet, this is the very essence of neoclassical microeconomics and the reader should never be allowed to confuse reactions to changes with spontaneous actions of the entrepreneurial type.

In a remark about Popper's reference to the 'logic of the situation' explaining the movements of a man crossing the street and dodging the cars,[1] Latsis exclaims: 'But entrepreneurial behaviour is usually very different from street crossing. For crossing the street we need little imagination, skill and deliberation. For entrepreneurial activity one often needs a high degree of imagination and skill; and these are processes which can hardly be reflected by crude single-exit models' (p. 224). Latsis recognises that simple models such as that of the profit-maximising firm 'have little application to actual entrepreneurial decision making' (p. 229), but he fails to tell the reader that these models were designed to explain and predict adjustments of an interdependences among market variables, not innovative, entrepreneurial activities.

The confusion between *reactions* (or actions) and the *effects* of reactions (or actions) has led many economists to aim at the wrong target. We are not primarily interested in businessmen, business decisions, business routine and business reactions; we are interested in firms only because of the effects of their actions and reactions upon production, resource use, incomes, and prices. Our main purpose is to explain and predict changes in these variables and we must look into the actions and reactions of firms only to the extent necessary for this purpose. Often this 'look' will be merely one at a mental construct (ideal type) of a firm rather than at a collection of real people in a corporate organisation. Yet, even if in specific situations the 'effects' cannot be explained without empirical research in the real people that make the particular decisions, these decisions are still only theoretical links in the chain that connects certain changes in conditions to certain changes in the variables to be explained. When Latsis offers a very interesting discussion of the 'bridge role' of certain psychological-physiological processes in the decision process of the individual and distinguishes stimulus from arousal and arousal from response (pp. 226–7) we ought to remind ourselves of the fact that even the response has only a 'bridge role' in the explanation or prediction of the variables in which we are ultimately interested: prices, incomes, allocations, and outputs.

---

[1] Popper [1945], *volume 2*, p. 97. My own illustration of the logic of the situation was that of an automobile overtaking a lorry on a highway. Machlup [1946], pp. 534–6; [1952], pp. 44–46, and [1967], pp. 6–7.

Latsis is fully aware of the widespread confusion between ***real people*** and ***mental constructs*** of decision-makers (which I have called 'imaginary reactors'). His awareness finds its most explicit expression in the statement that 'the neoclassical view stresses the ***situation*** and turns the decision making agent into a cypher' (p. 233). I cannot tell from the context whether Latsis sees this, as I do, as an advantage or rather as a defect of the neoclassical approach.

I have expressed this idea repeatedly in my writings and could select from a good many formulations. Here is one that I find especially pungent:

> the firm in the model world of economic micro-theory ought not to call forth any irrelevant associations with firms in the real world. We know, of course, that there are firms in reality and that they have boards of directors and senior and junior executives, who do, with reference to hundreds of different products, a great many things—which are entirely irrelevant for the microtheoretical model. The fictitious firm of the model is a 'uni-brain,' an individual decision-unit that has nothing to do but adjust the output and the prices of one or two imaginary products to very simple imagined changes in data.[1]

5 *What Does the Firm 'Know'?*

Marginalists as well as behaviouralists have fretted much about the amount of knowledge possessed by the 'firm'. Some marginalists have been overly generous, endowing the imaginary firm of their creation with 'perfect knowledge'. Some behaviouralists, on the other hand, have worried about the incompleteness, inaccuracy, and uncertainty of the information available to the real decision-makers in the business organisations of the real world; and, of course, they charged the marginalists with making ridiculously unrealistic assumptions when they hypothesised that the firm of their model was maximising its profits on the basis of 'adequate' knowledge.

Latsis seems to accept this criticism: 'Situational determinism assumes that the decision maker knows his situation better than he is assumed to know it in the conditions usually investigated in decision theory' (p. 211). And he states that 'the decision the agent will reach depends . . . also on his ***decision process***, namely, the way in which he will process the incomplete information available to him' (p. 228).

Whether the imaginary firm in the marginalist model 'knows' more or less than the real people who run firms in the real world, and how perfect, complete, accurate and certain its knowledge is, depends on the purpose for which the theorist constructs his model. For some purposes—for example, for a general-equilibrium system in welfare economics in which the allocation of resources is to be optimal at least in the sense used by Pareto—the imaginary firm has to be endowed with full knowledge of its technological and organisational possibilities as well as its buying and selling opportunities in the markets. For other purposes —for example, to show the effects which changes in conditions (*i.e.* changes in any of the mentioned possibilities and opportunities) will have upon inputs, outputs, and prices in competitive industries—the knowledge required is minimal: namely, just of the ***changes*** in conditions, not of the conditions themselves. (This is surely much less knowledge than any real person who can read the newspapers is apt to have.) No matter how incomplete, inaccurate, and uncertain the firm's knowledge is of its costs, revenues, profits, *etc.*, all that this firm has to know is the direction in which some pertinent magnitude has changed. 'What can be "imperfect"', I wrote in 1967 (p. 25) 'about the information on, say, a tax increase? Why should it take special theories of bureaucracy to explain how the news of a wage increase "flows" through various hierarchical levels up or down or across? Yet this, and this alone, is the information that is essentially involved in the theory of prices and allocation, since it is the ***adjustment to such changes in conditions*** for which the postulate of maximising behaviour is employed'.

---

[1] Machlup [1963], p. 133.

Alfred Schutz likened the ideal types constructed by social scientists to puppets acting out the roles assigned to them by their constructors. The firm in competitive-price theory is one of these puppets. 'By the grace of its constructor, he [the puppet] is endowed with just that kind of knowledge he needs to perform the job for the sake of which he was brought into the scientific world.'[1] The job of the ideal type called firm may be different from problem to problem; 'we must indicate the reference scheme within which the ideal type may be utilised, that is, the problem for the sake of which the type has been constructed'.[2]

6 *How Rational is the Firm?*

The quoted formulations by Alfred Schutz come from his essay on 'The Problem of Rationality in the Social World' and might be thought of as an appropriate transition to the question how much 'rationality' one may attribute to the firm, that is, to the imaginary reactor in price theory. Most discussions of the meaning and significance of rationality in economic theory would benefit from a study of Schutz's work. Writers might learn to be more explicit about what they mean, particularly whether they refer to reasonable actions, deliberate actions (routine, use of reliable insights, anticipations of desired ends, choices among alternative means for particular ends, but also choices among alternative ends), planned actions (projects), predictable actions, 'logical' actions[3] (as contrasted with traditional and habitual ones), to mention some of the distinctions discussed by Schutz.

With regard to the theory of the firm we ought to be told whether rationality refers to consistency with the information available to the actor (or reactor) or to consistency with the objectives and the preference scheme of the actor—or perhaps even to consistency with information and insights available not to the actor but to the observer. Latsis offers an interesting discussion of pronouncements on rationality by Karl Popper and J. W. N. Watkins and develops some interesting extensions of Watkins' ideas on imperfect rationality. What Latsis has to say about the separateness of 'appraisal of the situation', 'decision to act', and 'action' (execution of the decision) deserves attention. However, it remains to be shown that the suggested analysis of the decision process, its breakdown into several constituent phases, can be of substantial help in economic theorising. I do not question its significance in psychology, on the one hand, and in organisation theory and management science, on the other. That in 'situational determinism' the 'so-called rationality principle is absorbed into the postulates characterising the *homo economicus*' is fully recognised by Latsis (p. 229). I believe that this is all to the good, but I am afraid that Latsis thinks otherwise.

---

[1] Schutz [1943]; reproduced in his [1964], p. 83.

[2] *Ibid.* Perhaps I may in this context refer to an essay of mine in which I illustrated the use of seven different characterological and habitual ideal types and contrasted them with the *homo economicus*, an ideal type 'designed for interpreting observed consequences of men's actions' (Machlup [1970], pp. 135–9).

[3] This is Pareto's term.

7 *How Determinate is the Theory?*

This brings me to the question whether Latsis is justified in calling neoclassical microeconomics situational *determinism* in contrast to a programme of economic behaviouralism that he evidently does *not* regard as deterministic. We must ask in what sense behavioural theories of business firms can be regarded as non-deterministic.

Latsis finds that Herbert Simon committed a 'philosophical error' when he mistook 'situational determinism' for apriorism and 'psychologically observation-motivated theories for inductive ascent' (p. 229). As I see it, the fact that a model is constructed with a 'hard core' made out of fundamental hypotheses or postulates whereas other models are constructed with hypotheses derived from low-level behavioural generalisations on the basis of empirical observations should not make an essential difference in the application of the models. Any theoretical model—in contradistinction to a probabilistic model—, no matter how the built-in 'assumptions' were obtained, serves to be used in conjunction with some additional proposition—usually concerning an event or a change in conditions—so that a conclusion can be deduced from the conjunction. Both the additional proposition, which I once designated as the 'assumed change' (cause), and the conclusion, which I called the 'deduced change' (effect), are in the domain of construction but have counterparts in the domain of observation so that the correspondence can be tested.[1] In either case, the deduced change, the theoretical outcome of the whole business, is completely determined.

To put this statement in a slightly different form, any model designed to present (exhibit) a causal connection between an independent variable and a dependent variable under given conditions—conditions which may include fundamental hypotheses and a set of less fundamental assumptions on various levels of generality or specificity—must display the dependent variable as a logical consequence of all the premisses in the model. Since these premisses *imply* the conclusion, there can be no doubt about the *complete determinateness* of the result. Of course, whether this whole apparatus with its input and its determinate output applies to many situations of the real world, or to only a few, or to none at all, is a different matter. Determinate is only the theoretical conclusion inferred from the premisses. Nobody, however, can be certain whether the apparatus, the input, and the output will fit anything in the real world and will thus *satisfactorily explain* and *approximately predict* some observable outcome of the conjunction of an observed event with actual conditions prevailing in reality. Thus, the applicability of the model with its determinate conclusion is always open to question; it is *never* certain, no matter in how many instances it has been applied with success.

In one of my earlier essays—on 'Ideal Types, Reality, and Construction'[2]—I

[1] Machlup [1955], p. 13.

[2] Machlup [1960–61].

tried to show that mental constructs, ideal types and theoretical models designed for causal explanation are tools which *may* be useful regardless of the origin of what has gone into their construction. The ingredients or constituent parts may be abstractions or generalisations derived from empirical research, from case studies or statistical computations, but they may also be speculative idealisations, heroic hypotheses and grandiose postulates, spun by an inventive mind. Philosophers with ultra-empiricist bias will always want to inspect the certificates of origin of the assumptions imported into a theoretical model and will call for barriers against the import of assumptions not certified as the products of an empirical workshop. Such prejudices will neither make a successful model useless nor bestow pragmatic usefulness onto a model that fails to do a good job of explaining and predicting. The worth of a conceptual scheme does not depend on the descent of its constituent concepts.[1]

8 *The Wider the Choice the Harder the Choosing*

The basic argument for the behavioural theory of the firm is that decisions and actions of firms in reality depend on a variety of things, including the characterological make-up of their executives and the hierarchical set-up in their organisation and coordination. In order to explain and predict the actions (including the reactions to specified events) of a particular firm, we need a wide choice of models and have to select the right model that fits the type of firm. Assuming again that 57 models will cover most, though of course not all, of the types, we can hope that one of these models will do the job. Perhaps the specifications and directions will be sufficiently precise to enable us to eliminate a major part of the collection and thus narrow the choice to eight or ten models. While each of the models can furnish a perfectly determinate theoretical result, we can never be sure which of them is the appropriate one for the case in question. This is associated with the 'loss in generality' that the behavioural theorist had to admit. The increase in specificity increases the chance of accurate results provided we pick the right model for the case before us, but it reduces the chance of our finding that model.

Even if many of the models on the shelf had been constructed after painstaking observation of the decision-making processes of certain firms, one could never be sure that they would fit the same firms at a later time. Changes in the management personnel, in the division of responsibilities, in outside influences on the decision-making people, in governmental policies (say, with regard to anti-monopoly prosecution), and many other changes, internal as well as external, may have made the model 'constructed from life' quite obsolete and, hence, entirely unreliable in subsequent applications. The model is still in excellent working order, producing fully determinate theoretical outcomes, only that the produced (deduced) solutions would now fail to explain and predict the observed actions of the 'same' firm—if its sameness is defined only in legal terms. For the analyst the firm is no longer the same.

We may now have 57 determinate theoretical solutions but may be at a loss deciding which might be the right one for a concrete case in question. The probability of a good fit is not any greater and may be smaller than if we had used a very general, deliberately unrealistic model. In saying this, I may give the impression that the general model of the profit-maximising firm is recommended here as one of the alternatives in the task of explaining and predicting actions of particular firms. It is not; this model is used properly only for the effects of reactions of imaginary firms.[1]

---

[1] *Ibid.* pp. 55–7. I understand that most neopositivists would agree with my denunciation of these prejudices. Unfortunately, however, some of the prejudiced writers invoke the authority of (probably misunderstood) pronouncements by neopositivist philosophers to support their rejection of 'unrealistic' assumptions.

9 *Personal Types and Course-of-Action Types*

Many of the things suggested in the preceding pages can be said much better in the language of Alfred Schutz, or indeed in his words, except that his terminology and conceptual framework, derived from Max Weber on the one hand and Edmund Husserl on the other, have been resented and resisted in certain quarters. I am convinced that learning Schutzian language could help many in sorting out some messy notions in the methodology of the social sciences.

The social scientist, in order to grasp and interpret the 'other self', that is, people, their actions and reactions, and the effects of these actions and reactions, has to form 'a construct of a typical way of behaviour, a typical pattern of underlying motives, or typical attitudes of a personality type.[2] Such types can be

> arranged according to the degree of increasing anonymity of the relationship among contemporaries involved and therewith of the construct needed to grasp the other and his behaviour. It becomes apparent that an increase in anonymity involves a decrease in fullness of content. The more anonymous the typifying construct is the more detached is it from the uniqueness of the individual fellowman involved, the fewer aspects also of his personality and behaviour pattern enter the typification as being relevant for the purpose at hand, for the sake of which the type has been constructed. If we distinguish between (subjective) personal types and (objective) course-of-action types we may say that increasing anonymisation of the construct leads to the superseding of the former by the latter. In complete anonymisation the individuals are supposed to be interchangeable, and the course-of-action type refers to the behaviour of 'whomsoever' acting in the way defined by the construct.[3]

The social scientist, the 'scientific observer of social interaction patterns', constructs

> typical course-of-action patterns corresponding to the observed events. Thereupon he co-ordinates to these typical course-of-action patterns a personal type, namely, a model of an actor whom he imagines as being gifted with consciousness. Yet it is a consciousness restricted to containing nothing but all the elements relevant to the performance of the course-of-action patterns under observation and relevant, therewith, to the scientist's problem under scrutiny.[4]

---

[1] Still, if I were given the assignment to interpret in a concrete case the actions of a particular firm and were perplexed by having to find the correct model from among a wide choice of behaviourally specific models, I would be tempted to try first that unduly unrealistic, simplistic, anonymous general model and see what it could do for me.

[2] Schutz [1853], p. 12; [1962], p. 17.

[3] Schutz [1953], p. 13; [1962], pp. 17–18.

[4] Schutz [1953], pp. 31–32; [1962], p. 40.

> Yet these models of actors are not human beings living within their biographical situation in the social world of everyday life. Strictly speaking, they do not have any biography or any history, and the situation into which they are placed is not a situation defined by them but defined by their creator, the social scientist. He has created these puppets or homunculi to manipulate them for his purpose. A merely specious consciousness is imputed to them by the scientist which is constructed in such a way that its presupposed stock of knowledge at hand (including the ascribed set of invariant motives) would make actions originating therefrom subjectively understandable, provided that these actions were performed by real actors within the social world. But the puppet and his artificial consciousness is not subjected to the ontological conditions of human beings. The homunculus was not born, he does not grow up, and he will not die. He has no hopes and no fears: he does not know anxiety as the chief motive of all his deeds. He is not free in the sense that his acting could transgress the limits his creator, the social scientist, has predetermined. He cannot, therefore, have other conflicts of interests and motives than those the social scientist has imputed to him. He cannot err, if to err is not his typical destiny. He cannot choose, except among the alternatives the social scientist has put before him as standing to his choice.[1]

The methodological principles guiding the construction of ideal types and models and their uses for, or applications to, particular problems are the same for types of different degrees of anonymity. It is a mistake to think that models containing types of greater anonymity yield fully determinate results while models with less anonymous types yield less determinate results. Any well-constructed model furnishes precisely the results the constructor wants it to furnish. Thus the most useful model has only a 'single exit', if I may use Latsis's expression. And this holds for a highly 'behavioural' model no less than for one with only fully anonymous, single-minded maximisers of profit. Indeed, if the postulate of profit maximisation in certain circumstances—*e.g.* in cases of oligopoly—fails to make the model produce completely determinate results, we regard it as unsuitable for instances of this sort and try to replace it with one that yields just the results that we want it to deliver.

10 *Behavioural Determinism*

If every good theoretical model yields just one solution—because we want to work only with models that furnish determinate solutions—but if some will do so with a simple, most general objective function while others need more specific 'psychological' and 'organisational' behaviour functions, Latsis may have to distinguish two kinds of determinism: situational determinism and behavioural determinism.

I would not choose these terms, however, because I think that the adjectives do not bring out the real difference and the noun is misleading in several respects. The problem of determinism should, in my opinion, not be brought into the discussion of model building and theory formation in the theory of the firm or in economics in general, except where probabilistic models are used. However, if

---

[1] Schutz [1953], p. 32; [1962], p. 41.

Latsis insists that an explanatory scheme which fully explains a specified change as an inevitable effect of a conjunction of a specified cause with specified conditions should be designated as an instance of determinism, I am willing to humour him, provided he recognises that his behavioural approach will become satisfactory only as it becomes just as 'deterministic' as the situational approach.

Neither Latsis nor the proponents of the behavioural programme have, if I understood them correctly, meant to introduce probabilistic models into the theory of the firm. They have merely proposed to distil from the findings of empirical observation of actual decision processes a bevy of behavioural types for use in empirically richer models. I conclude that the disanonymisation of the fundamental postulate of economic behaviour into a variety of particular types and subtypes of behaviour does not justify distinguishing the two research programmes by awarding to one and refusing to the other the designation 'deterministic'.

FRITZ MACHLUP
New York University

REFERENCES

BAUMOL, W. J. [1959]: *Business Behaviour, Value, and Growth.*

BAUMOL, W. J. [1962]: 'On the Theory of the Expansion of the Firm', *American Economic Review*, **52**, pp. 1078–87.

COHEN, K. [1960]: 'Simulation of the Firm', *American Economic Review*, **50**, pp. 534–40.

CYERT, R. M. and MARCH, J. G. [1963]: *Behavioural Theory of the Firm.*

LATSIS, S. J. [1972]: 'Situational Determinism in Economics', *British Journal for the Philosophy of Science*, **23**, pp. 207–45.

MACHLUP, F. [1939]: 'Evaluation of the Practical Significance of the Theory of Monopolistic Competition', *American Economic Review*, **29**, pp. 227–36.

MACHLUP, F. [1946]: 'Marginal Analysis and Empirical Research', *American Economic Review*, **36**, pp. 519–54.

MACHLUP, F. [1952]: *The Economics of Sellers' Competition.*

MACHLUP, F. [1955]: 'The Problem of Verification in Economics', *Southern Economic Journal*, **22**, pp. 1–21.

MACHLUP, F. [1960–1]: 'Idealtypus, Wirklichkeit und Konstruktion', *Ordo*, **12**, pp. 21–57.

MACHLUP, F. [1963]: *Essays in Economic Semantics.*

MACHLUP, F. [1967]: 'Theories of the Firm: Marginalist, Behavioural, Managerial', *American Economic Review*, **57**, pp. 1–33.

MACHLUP, F. [1969]: 'If Matter Could Talk', in S. Morgenbesser, P. Suppes, and M. White (*eds.*): *Philosophy, Science, and Method*, pp. 286–305.

MACHLUP, F. [1970]: 'Homo Oeconomicus and His Class Mates', in M. Natanson (*ed.*): *Phenomenology and Social Reality*, pp. 122–39.

POPPER, K. R. [1945]: *The Open Society and Its Enemies.*

SCHUTZ, A. [1943]; 'The Problem of Rationality in the Social World', *Economica*, N.S., **10**, pp. 130–49.

SCHUTZ, A. [1953]: 'Common-Sense and Scientific Interpretation of Human Action', *Philosophy and Phenomenological Research*, **14**, pp. 1–38.

SCHUTZ, A. [1962]: *Collected Papers*, **1**, *The Problem of Social Reality.*

SCHUTZ, A. [1964]: *Collected Papers*, **2**, *Studies in Social Theory.*

WATKINS, J. W. N. [1970]: 'Imperfect Rationality', in R. Borger and F. Cioffi (*eds.*): *Explanations in the Behavioural Sciences*, pp. 167–217.

WILLIAMSON, O. E. [1964]: *Economics of Discretionary Behaviour: Managerial Objectives in a Theory of the Firm.*

SITUATIONAL DETERMINISM IN ECONOMICS: THE IMPLICATIONS OF LASTIS'S ARGUMENT FOR THE HISTORIAN OF ECONOMICS

Lastis's paper is an important and constructive contribution to the methodology and history of economics even if one questions his proposition that 'the adoption of antipsychologism as a heuristic canon is not only unnecessary but, by restricting permissible types of explanatory generalisation, may halt progress in microeconomics' (Lastis [1972], p. 229). For, as befits a philosopher of science, his method is more important than his implied historical judgment. By focusing attention on a fairly narrow, long-established, and theoretically sophisticated segment of economic analysis, and providing a more penetrating and methodologically more precise review of its evolution, he has helped to elucidate the nature and significance of scientific 'progress' in economics. In the process he has demonstrated that Lakatos's 'research programme' concept is a more powerful weapon than Kuhn's blunt and multi-faceted 'paradigm' instrument; and, after the initial dust of controversy has settled, he may be seen to have provided a more secure basis for agreement and mutual understanding between economists of differing methodological persuasions.[1]

Within the scientific community, as among members of the general public, economics has long been regarded as the most advanced, prestigious, and scientific of the social sciences. This reputation derives from a familiar combination of features, including the technical complexity and manifest social significance of its subject-matter, and the sophistication of its mathematical and quantitative methods. But most important of all, at least to the community of scholars and scientists, is the remarkable fertility and continuity of its central corpus of theory. Of course, a cynic might ask whether this continuity is evidence of anything other than sheer survival value, a question which fortunately need not be considered here. However it should be noted that at least one distinguished recent philosopher has maintained that the definition of 'science' is ultimately reducible to a matter of 'tradition'.[2] And the contention is relevant, since Lastis's paper provides clear evidence of doctrinal continuity in the orthodox microeconomic theory of the firm. This continuity has been maintained in a number of ways, for example by successive theoretical modifications, refinements, and reinterpretations, and by determined resistance to the numerous critics who have proposed to reformulate the inherited theory drastically, abandon it, or apply it to problems which it was incapable of handling.

In the process, the theory has become more systematic, more rigorous, and more restricted in scope; and those who are disturbed by these developments may be consoled by the suggestion that 'the exact sciences have won some of their earlier successes by avoiding the kind of problems still too difficult for their yet undeveloped powers'.[3] Whether such gains are more than offset by the losses entailed in neglecting the more difficult problems is obviously a matter of historical judgment, one that will depend on the selection of appropriate criteria of scientific, or perhaps social, importance. There is obviously a danger of undue concentration on trivial 'puzzles' while major gaps or anomalies are ignored.

---

[1] This may occur as the specific strengths and limitations of different 'research programmes' become more widely appreciated. For an earlier constructive effort to expose the fundamental irreconcilability of different approaches to economics see Briefs [1961].

[2] Gallie [1957].

[3] Deutsch [1958], p. 28.

Moreover any judgment will necessarily depend on the time perspective adopted, since short run losses may eventually be outweighed by the higher long run returns from scientifically more productive 'roundabout processes', as in Austrian capital theory. Finally, we must not only consider the effects on the theory itself or the segment of the discipline in question, but also the effects on the discipline as a whole, and on related disciplines. Thus even if we convert Lastis's careful conjecture into the counter-factual proposition that 'the adoption of anti-psychologism as a heuristic canon' has in fact halted 'progress in economics', it is at least arguable that the economists' wilful rejection of the more intractable aspects of the theory of the firm may have contributed indirectly to the adoption of alternative strategies by other scholars or by lapsed economists: for example organisation theory, and studies of managerial decision-making and administrative behaviour.

Needless to say, Lastis's account of the evolution of situational determinism in the theory of the firm will not satisfy the historian of economics, since he has deliberately abstracted from the historical context in which it occurred. Nevertheless, his approach is suggestive, as may be indicated by a brief sketch of similar, but earlier, developments in the theory of consumer's behaviour (to which he alludes in a footnote).[1] The protracted controversy surrounding this subject can be traced back through the *Methodenstreit* of the 1880s to the early nineteenth century attacks on classical economics by literary and social reformist critics, who maintained that the economists had grossly overrated the importance of self-interest and thereby presented a debased and one-sided view of human nature and society. The later phase of criticism, which gathered momentum from the 1880s, rested on more secure foundations as the growth of experimental psychology encouraged the repudiation of the hedonistic theory of behaviour, which was presumed to underly the new subjective (or so-called 'psychological') theory of value—especially the marginal utility version propounded by W. S. Jevons. In the ensuing debate among the economists those who contended that economics was vitiated, or at least seriously impaired, by its dependence on unsound psychological foundations looked for support, successively, to the physiological and biological theory of William James, the instinct theory of William MacDougall and the behaviourism of John Broadus Watson. The issue had to be faced squarely for MacDougall, in particular, threw down the gauntlet, by declaring

> it would be a libel, not altogether devoid of truth, to say that classical political economy was a tissue of false conclusions drawn from false psychological assumptions.... The great assumption of the classical political economy was that man is a reasonable being who always intelligently seeks his own good and is guided in all his activities by enlightened self-interest, ... [whereas in fact] mankind is only a little bit reasonable and to a great extent very unintelligently moved in quite unreasonable ways.[2]

---

[1] P. 208, footnote 3. The following account will be confined to the American discussion, which was intense and protracted from the early 1890s to the later 1920s. It is based on Chapter 3 of my unpublished Ph.D. thesis, 'Methodological Controversy as an Approach to the History of American Economics, 1885–1930', Johns Hopkins, 1953. However, many of the same issues were raised in Britain albeit less heatedly and persistently, for example by such heterodox writers as J. A. Hobson and P. Sargent Florence.

[2] MacDougall [1908], p. 11.

Among those American economists disposed to plead guilty to MacDougall's accusations were such significant figures as Thorstein Veblen, Wesley Mitchell, Frank A. Fetter, and John Maurice Clark. Individually or collectively they maintained that in its existing condition economic theory was unrealistic, incomplete, and incapable of providing a basis for predictions since it had adopted an unscientific subjective approach, underestimated the importance of irrational impulses and habitual conduct, and neglected societal influences on individual behaviour. It was therefore unable to explain the origin and nature of consumer's wants, individual responses to the availability of new commodities, or the effects of advertising and social emulation. Of course, these admissions did not go unchallenged among the leading defenders of the orthodox tradition, among whom were such powerful spokesmen as Irving Fisher, Arthur T. Hadley, Herbert J. Davenport, and Frank H. Knight, who either denied the relevance of psychology to economics or regarded it as a peripheral modifier rather than a decisive negation of the subjective theory of value.

In terms of Lakatos's methodology of research programmes it is especially interesting to trace, over a period of three or four decades, the gradual emergence of a concensus of opinion that in its central theory of value economics required no psychological theory, but only a fundamental postulate—variously termed the principal of rationality, or the logic of choice. The routes by which this conclusion was reached were necessarily various. Fisher, for example, from his *Mathematical Investigations in the Theory of Value and Prices* (1892) through his *Elementary Principles of Economics* (1912) consistently maintained that psychology was irrelevant to economics. Davenport, after stressing the importance of psychology in the process of valuation in his *Value and Distribution* (1908), subsequently produced a textbook, *The Economics of Enterprise* (1913), which wholly ignored valuations and defined economics as 'the science which treats phenomena from the standpoint of price' (p. 25). Mitchell, after several ambitious programmatic statements of the importance of psychology eventually settled for a Veblenian type of 'pecuniary logic', after a stimulating discussion of the influence of money on individual behaviour.[1] Knight, in characteristically iconoclastic fashion, embraced behaviourism as an adequate foundation not on account of its positive merits, but because it was 'less a psychology, than a dogmatic repudiation of everything properly to be called by that name'.[2] In Knight's hands the separation of economics and psychology became absolute, for science was to him 'the technique of prediction'[3] and

> The basis of a *science* of conduct must be fixed principles of action, enduring and stable motives. It is doubtful, however, whether this is fundamentally the character of human life.[4]

Anyone who interpreted the premises of economics as a literal description of human behaviour was badly mistaken, but the postulate of rationality, like other scientific premises, was not actually true—in other words, it was an 'irrefutable core'. The critics had

---

[1] *Cf.* his [1916] and the title essay in his [1937]. His main works on business cycles were almost purely quantitative, and owed nothing to psychological theory.

[2] Knight [1931], p. 64. See also his [1921*a*], pp. 203–4, 208.

[3] Knight [1924]; p. 232.

[4] Knight [1921*a*], p. 53, *italics in original.*

> made little headway in bringing forth substitute principles. I do not believe that they ever will. The structures are valid as *limitations* on the familiar reasonings, not as negations. The principles of the established economics are partial statements, but sound as far as they go, and they go about as far as general principles can be carried.[1]

And there, at least as far as the majority of American economists were concerned, the matter rested.[2]

Moderate commentators, like Zenas Clark Dickinson,[3] fully appreciated that there were important problems on the borderlines between economics and related disciplines. But, like the majority of his professional colleagues, he was content to leave them to specialists in other fields. The scope of economics had been redefined, on traditional lines, by discussion and eventual agreement. In the process the basic theory had been explicated, simplified and effectively protected from refutation.

Whether such a process was inevitable or desirable is, of course, quite another matter.

A. W. COATS
University of Nottingham

REFERENCES

BRIEFS, H. W. [1961]: *Three Views of Method in Economics.*

DEUTSCH, K. W. [1958]: 'Scientific and Humanistic Knowledge in the Growth of Civilisation', in Harcourt Brown (*ed.*): *Science and Creative Spirit, Essays in Humanistic Aspects of Science.*

DICKINSON, Z. C. [1922]: *Economic Motives: A Study in the Psychological Foundations of Economic Theory, with some Reference to Other Social Studies.*

GALLIE, W. B. [1957]: 'What Makes a Subject Scientific?', *British Journal for the Philosophy of Science*, **8**, pp. 118–39.

KNIGHT, F. H. [1921a]: *Risk, Uncertainty and Profit.*

KNIGHT, F. H. [1921b]: 'Traditional Economic Theory: Discussion', *American Economic Review*, **11.**

KNIGHT, F. H. [1924]: 'The Limitations of Scientific Method in Economics', in R. G. Tugwell (*ed.*): *The Trend of Economics.*

KNIGHT, F. H. [1931]: 'The Relation of Utility Theory to Economic Method in the Work of William Stanley Jevons and Others', in Stuart A. Rice (*ed.*): *Methods in Social Science: a Case Book.*

LASTIS, S. J. [1972]: 'Situational Determinism in Economics', *British Journal for the Philosophy of Science*, **23**, pp. 207–245.

MACDOUGALL, W. [1908]: *An Introduction to Social Psychology.*

MITCHELL, W. C. [1916]: 'The Role of Money in Economic Theory', *American Economic Review*, **6.**

MITCHELL, W. C. [1937]: *The Backward Art of Spending Money.*

---

[1] Knight [1921*b*], p. 145.

[2] There has been, of course, much subsequent discussion of ordinal and cardinal utility, indifference curve analysis, revealed preference, etc. But these topics have not entailed any serious questioning of the principle of rationality or the fundamental relationship between economics and psychology. It should, however, be noted that in the period under review there was a considerable effort among American economists to develop a 'social' approach to value and welfare theory, in opposition to the orthodox tradition of methodological individualism. However, it would be inappropriate to raise these matters here. They are considered, for the pre-1930 period, in my thesis cited *above*, p. 286, footnote 1.

[3] Dickinson [1922]. This is still the best published review of the debate up to that point.

A CORRECTION CONCERNING COMPLEXITY

In an earlier paper[1] I defined the complexity of a proposition H as equal to the amount of information in it, that is, as $I(H) = -\log P(H)$. It disturbed me that the proposition $0 = 1$, which looks simple, is thus attributed infinite complexity, but I accepted this implication because this proposition implies all other propositions. K. S. Friedman (private communication) has now pointed out an objection to my postulate that $H . K$ is more complex than H. He shows that it leads quickly to the conclusion that $A \vee B$ is simpler than A, which is counter-intuitive. The postulate would be reasonable if H and K were entirely independent of each other (in fact it is reasonable to assume that the complexities are additive in this case), but I must now withdraw the definition of complexity as given. Fortunately this has very little affect on the rest of the paper. The definition might make sense if restricted to a conjunction of 'atomic propositions', or if $P(H)$ is interpreted not as that of the proposition H, but as the maximum probability, ***regarded as a linguistic text,*** of an expression of H. This probability could be evaluated approximately by regarding the language as a Markov chain of any finite order. Such approximations are familiar to cryptanalysts and are mentioned, for example, in Shannon's basic paper on communication theory. For more details, see my forthcoming paper 'Explicativity, corroboration, and the relative odds of hypotheses', in the proceedings of the conference on 'Methodologies: Bayesian and Popperian', Columbia, South Carolina, November 1973.

As mentioned in my paper, Valéry (1921) described a figure as 'geometric' if it can be traced by motions which can be expressed in few words. Perhaps he obtained this idea from Émile Lemoine's book *Géométrographie*[2] where measures of complexity of geometrical constructions are given. See also J. L. Coolidge, *A Treatise on the Circle and Sphere.*[3]

I. J. GOOD
Virginia Polytechnic Institute and
State University

REFERENCES

COOLIDGE, J. L. (1961): *A Treatise on the Circle and the Sphere.*

GOOD, I. J. (1968): 'Corroboration, Explanation, Evolving Probability, Simplicity and a Sharpened Razor,' *British Journal for the Philosophy of Science*, **19**, pp. 123–43.

LEMOINE, E. (1902): *Géométrographie.*

VALÉRY, P. A. (1921): 'Eupalinos ou l'Architecte-Dialogue des Morts', *La Nouvelle Revue Française*, 90, pp. 237–85.

---

[1] Good [1968].
[2] Lemoine [1902].
[3] Coolidge [1916], p. 170.

Brit. J. Phil. Sci. 25 (1974), 291–300 *Printed in Great Britain*

Reviews

RESCHER, N. [1973]: *The Coherence Theory of Truth.* Oxford: Oxford University Press. £5.50. Pp. xiv + 374.

The author starts by acknowledging that the correspondence theory of truth is perfectly adequate if we are only concerned with the 'meaning' or 'definition' of truth; but he emphasises its inadequacy as a criterion of truth (see for example p.24). The coherence theory, he undertakes to show, is best construed as providing a criterion of factual or empirical truth. And, he insists, this is well in line with the idealist tradition. Now whether this be so or not, Rescher's arguments for this continuity are strangely unconvincing, there being a most marked change of mood on page 39 when he begins to formulate the strategy of his approach to coherence.

Roughly speaking, the problem is seen like this. In science we frequently find ourselves faced with an inconsistent set of sentences, from which, somehow, we wish to distil a consistent theory. A typical case would be when a prediction of some scientific theory is refuted. Logic only tells us that some segment of the aggregation of scientific theory and auxiliary hypotheses (if any) needs correction; it does not say which segment. Our aim is clearly to salvage as much truth as we can from the inconsistent wreckage. The trouble is that every conceivable consistent theory lurks somewhere in the ruins, so that we seem to be equipped with no clues at all. This is where Rescher's criterion is supposed to start twitching. And though, the author assures us, it is not a 'guaranteeing' criterion that by necessity leads to the truth, it is an 'authorizing' one (pp. 16–22); in other words, it provides us with a 'rational warrant' for the claim that we have acquired not only something true, but as much truth as we are jolly well entitled to.

The logic of the criterion is agreeably simple. Let **S** be an inconsistent set of sentences, what Rescher calls 'data' or 'truth-candidates' (pp. 53–64). **S** is assumed finite. These data cannot of course all be true, and indeed it is not assumed of any one that it is. All that is assumed is that the truth lies somewhere within **S** (p. 72). By 'taking a proposition to be a datum we propose to class it as true *whenever possible*' (p. 55). Given **S** our task is to sort out 'those propositions among the totality of relevant data that exhibit the best fit with the rest —that "maximally cohere with the data" ' (p. 74). Consider, therefore, says Rescher, all the maximal consistent subsets of **S**; that is, the strongest theories to be obtained from **S** by trimming it down to consistency. Any sentence that follows from all these consistent subsets is known as an 'inevitable consequence'. Since such consequences will be procurable however we choose to thin out **S**, we seem to have some sort of reason for holding them to be true. What is slightly surprising is that they may not all be tautologies. This is far and away the most interesting idea in the book.

As an illustration (see p. 83) consider the set $\mathbf{S} = \{p, p \supset q, \sim q\}$, where p and q are atomic sentences. The three maximal consistent subsets are $\{p, p \supset q\}$, $\{p, \sim q\}$, and $\{p \supset q, \sim q\}$. Thus $q \supset p$ is an inevitable consequence. Rescher's criterion would certainly, therefore, indicate it as true. We have, as it were, generated truth 'not *ex nihilo* (which would be impossible) but yet from a basis that does not itself demand any prior determinations of truthfulness as such' (p. 70); 'a set of data that contains no [acknowledged] truths can yield truths' (p. 92). But this will not always be the case, as is easily seen by adjoining the sentence $\sim (q \supset p)$ to **S**. All inevitable consequences of the new set will be tautological.

Under what conditions does **S** have inevitable consequences that are contingent? Rescher never addresses himself to the question. Indeed, hardly anywhere in the book does he attempt to prove any general theorems, being content to work through the medium (and, I am afraid, tedium) of examples. This is a pity. For it takes away most, if not all, of the point of formalisation. But in fact Rescher is not much interested in confining himself to inevitable consequences; 'this mooted conception of "consequences" is too restrictive', he asserts (p. 82), presumably because the truth-yield is far from comprehensive. At which he considers a variety of ways in which we might come to 'prefer' some of the maximal consistent subsets, so that the 'truth' would be the common consequences of all these preferred theories (Chapter 5). This is the coherence criterion of truth.

I shall return in a moment to these preference mechanisms. First, however, I wish to note a simply monumental objection to this proposed criterion of truth—the objection that *it is not a criterion*, since it cannot always be applied, *even in principle*. The reason is one of the most celebrated results of mathematical logic, Church's Theorem, of which one could believe Rescher had never heard. *The determination of consistency in elementary logic is a non-effective task.* It is significant that almost every example provided by Rescher (and there are a good many, some repeated time and time again) is drawn from sentential calculus. This he claims to be a simplification of no importance (p. 86): 'In any such situation, the determination of maximal consistent subsets and the application of preferential machinery could be carried through in a way strictly parallel with the procedure in the propositional case.'

Thus in the general case there is simply no method by which the maximal consistent subsets can be sorted out; so no method for determining what is true. And there is no doubt at all that Rescher intends his criterion to be applicable in practice. He writes (p. 238), when discussing a variable criterion *C* (which could well be his own): 'All I ask is that we be *serious* about *C*; that *C* really and truly is the criterion we are actually to employ in practice for the determination of factual truth.' Indeed, it is essential that he adopt this attitude. For although an impracticable criterion of *acceptance* might be defensible on the grounds that it gives a 'rational reconstruction' of what we ought to do, this is not the case with a criterion of *truth*. For as Rescher is keen to concede, the correspondence theory does this. And Rescher makes plain at the outset (p. 16) that he is out to construct a criterion of truth, not a criterion of acceptance. (Admittedly, for example on p. 240, he seems to merge the two to a certain extent.)

However, let us suppose this objection overruled. Of the modes of preference amongst maximal consistent subsets that are sketched in Chapter 5 the one most used later relies on what Rescher calls 'plausibility indexing' (pp. 114–23). This seems to depend on the fact that usually some members of **S** will appear more plausible than others. Construct, says Rescher, a numerical scale to assess these plausibilities; for each maximal consistent subset of **S** compute the average plausibility of its elements; and prefer those subsets with the better average. (This frequently gives the result that we would have obtained by simply chopping out of **S** its most implausible element. See, for example, the amazing application of coherence analysis to the 'justification of induction' on pp. 225–30.) But since Rescher gives no indication of what the unit of plausibility is (and a change in this will affect the averaging procedure a good deal, since we have a zero), this method is quite obviously highly arbitrary. Indeed, the slapdash approach to plausibility indexing that we get in most examples (see, perhaps, p. 305) suggests that usually almost any result could be arranged by a judicious use of 'coherence analysis'. And the same thing goes, more or less, for the other canons of preference. No doubt this is how the author can see the whole of scientific method, amongst other things, falling within the scope of his criterion (p. 256).

Rescher acknowledges (p. 180) that plausibility considerations are 'extralogical'; but to grant this, he maintains, 'is definitely not to categorize [them] as haphazard, arbitrary, or devoid of rational warrant'. But to say *this*, or indeed anything else in the book, is definitely not to supply these considerations with any jot or scintilla of rational warrant. It is a pity that Rescher did not forgo the temptation to be extralogical, at least for a while; for as I have noted, he has produced a logical idea of some interest, and it would have been worth seeing what could be made of it. If, as I strongly suspect, the answer is nothing, then the same would have had to hold for any extralogical extension.

One worry which arises already with inevitable consequences is that, quite obviously, what the criterion throws up as true depends in a pretty crucial way on how the data are formulated. Rescher admits this (pp. 102, 111, 180 f.), but is blissfully unconcerned. 'When the issue of the truth of what is declared or maintained is before us, the mode of formulation is not a grammatical irrelevancy but a definitive aspect of the problem.' Now this seems to me too naturalistic by half, and even to conflict with some of the author's later proposals (p. 312). But even if it is not, surely we are entitled to some sort of explanation of why *trivial* reformulations of the data lead to *important* changes in what comes out as true? Of such an explanation there is no sign; and in my opinion no hope.

The latter part of the book makes many sweeping claims on behalf of the 'coherence criterion of truth'; that it helps us with the problem of induction, that it gives a new analysis of counterfactual conditionals (Chapter 11). There is also what is called a 'pragmatic justification' of the coherence theory. I need hardly say that these claims are quite unfounded.

The book is, of course, too long. But that does not mean that there are not important things omitted. In particular one must deplore Rescher's most casual approach to the previous literature.

If, given the set of data **S**, we take it for granted that the inevitable consequences of **S** are to be judged true, then we are in fact in the situation of having a set of exclusive and exhaustive alternatives amongst which to choose. For all maximal consistent subsets contradict one another; and taking the inevitable consequences as true is equivalent to taking one of these subsets as true. Now Rescher's 'coherence analysis' may occasionally permit us to select just one of these maximal consistent subsets as encapsulating the truth, but frequently several together will count as 'preferred'; in this case what we conclude to be true is in effect a disjunction of the axioms of the preferred subsets. In other words, the problem, and even the solution, have startling similarities to Levi's altogether more modest, and infinitely more effective book, *Gambling with Truth*. (Compare, for example, pp. 84 f. of Levi's book with the paragraph on pp. 93 f. of Rescher's.) Does Rescher attempt to compare Levi's work with his own? Indeed he does not. However, he twice mentions Levi's book, once citing him as an 'authority', so there can be no doubt that he realises that it covers similar topics.

In particular, both authors discuss Shackle's 'potential surprise', and show how it can be articulated within their own approaches. Rescher is severely critical of Hamblin's earlier attempt to do this (p. 351), but he does not mention Levi's. This is almost embarrassing when one considers that his own theory of plausibility appears to fall foul of almost the same criticisms that he directs at Hamblin. The latter is taken to task because his theory registers the plausibility of $P \vee \sim P$ as the greater of the plausibilities of P and of $\sim P$; thus either 'On the next toss this die will come up < 3' or 'On the next toss this die will come up > 3' is *maximally plausible* (p. 349). This is held to be absurd because, presumably, each of these sentences has 'fair-to-middling' plausibility. But now look at Rescher's theory. According to (P4) on p. 115, if P & Q is consistent (an unexplained requirement) then its plausibility index is the greater of the plausibility indices of P and Q. (For Rescher, note, a high plausibility is recognised in a low plausibility index, and *vice versa*; see p. 114.) So consider a roulette wheel; since 'Not *rouge* & not *noir*' is consistent, and since perfect symmetry obtains, its plausibility index will be the plausibility index of 'Not *rouge*'. But it is absurd that 'Zero' should have the same plausibility as 'Not *rouge*'.

Another person, incidentally, whose work might have been considered in this context is L. J. Cohen.

Here is another example. In Chapter 13 Rescher discusses the possibility of a non-foundationalist theory of knowledge. In a note on p. 318 we read: 'Popper's position is too complex for treatment here. . . . Popper's own falsificationist theory has from our standpoint a decidedly negative tendency in that it calls, in the final analysis, for abandoning in the factual domain any and all application of the traditional conception of truth. . . . A technical concept of "verisimilitude" is Popper's surrogate for the traditional concept of truth, but its ramifications are complex and require more space than the incidental discussion allows.' But why is the discussion 'incidental'? And where comes the idea that Popper substitutes verisimilitude for truth? This is quite extraordinary, since Popper actually defines verisimilitude in terms of truth; and Rescher himself says on page 198: 'Regarding this idea of "degrees of truth" as representing approximations to correspondentist truth see K. R. Popper . . .'. Yet in fact for Popper verisimilitude is no more 'applicable' than truth is; it too is just a matter of guesswork. (See *Conjectures and Refutations*, p. 234.)

Lastly, is not David Hume entitled to the credit—if credit there be—for this remark (p. 255)? 'In a real sense cognition is ancillary to practice and *feeling* becomes the arbiter of knowledge.'

DAVID MILLER
University of Warwick

SIMON, MICHAEL A. [1971]: *The Matter of Life. Philosophical Problems of Biology*. New Haven and London: Yale University Press. £3.75 cloth. Pp. xi+258.

The central problem with which Simon deals in this book is that of the distinctiveness of biology from physics: just how distinct *is* biology from physics and what generates this distinctiveness? If present-day biology, unlike chemistry, is largely independent of present-day physics, is it thereby less scientific?

With a view to answering these questions, Simon examines biology from a number of different aspects. In Chapter *1* he investigates biology's status as a science and its claim to autonomy. In Chapter *2* he discusses the logical forms of the types of explanations used in biology and their similarity or otherwise to those found in physics. Chapter *3* is concerned with the role of 'models' in biology and contains a detailed case-study of the development of the theory of the gene, as an example. In Chapter *4*, Simon ferrets out and discusses some tacit assumptions and background explanatory theories found in biology. Finally, Chapter *5*, entitled, 'The Extensions of Biological Thought', touches on that hazy area where biology borders on psychology and anthropology and also discusses the relations (or rather, lack of them) between biology and ethics and the notion of progress in evolution.

Simon's attempted solution of his central problem of the distinctiveness of biology consists of three main parts. He first argues that this distinctiveness does *not* lie in biology's not being a science. This claim is established in a negative fashion: Simon shows that a number of criteria which appear to exclude biology from the scientific realm (*e.g.* the criterion based on the requirement that all true sciences contain universal laws) in fact fail to exclude biology. Secondly, he argues that biology's distinctiveness does *not* lie in its using forms of explanation irreducibly different from those used in physics. For example, Simon cogently argues that functional explanations as used in biology do not rest on any presuppositions about design which would essentially separate biology from physics. Thirdly, Simon argues positively that biology's distinctiveness arises from the fact that its subject matter allows itself to be viewed (I think Simon would even say, *demands* to be viewed) simultaneously as a heap of molecules subject to the laws of physics and chemistry, *and* as a coherently organised whole, exhibiting purposiveness at all levels of organisation. This dual nature of living organisms generates what Simon dubs the 'complementarity' of the reductionist and anti-reductionist approaches within biology:

> The organism is at once an inscrutable man-like object [to be studied by anti-reductionist methods] and a soluble puzzle of intelligible science [to be studied by reductionist methods]. It is that tension that is responsible for the special character of the science of living things (p. 239).

According to Simon, in order to fully investigate the organism, biology has and *needs* two directions for its investigations: upwards, as advocated by the anti-reductionists, toward system laws which would be general laws governing all hierarchically organised systems; and downwards, as advocated by the reductionists, toward chemistry and physics. In fact, Simon appears to be claiming that even if the structure and behaviour of an organism were fully explicable and predictable in physico-chemical terms, this would not be enough. To achieve a 'total understanding' of the organism, it must also be viewed from the anti-reductionist *i.e.* 'holist' position.

While I found the first two (negative) parts of Simon's solution to his central problem clearly expressed and on the whole convincing, there seem to be weaknesses in this third (positive) part of his solution. These weaknesses can, I think, be traced back to the rather uncritical acceptance of a somewhat eclectic general philosophy of science, composed of elements drawn from various traditions.

For example, he accepts a subjectivist view of scientific explanation, defining a good explanation as one which confers 'intellectual satisfaction'. From Kuhn and Polanyi he accepts the view of scientific controversies as clashes of commitment, the two sides of which cannot be rationally appraised. The reductionist and anti-reductionist views, Simon claims, represent 'fundamentally different ways of viewing the world of living things' and scientists 'commit' themselves to the one or the other for 'personal, psychological or sociocultural' reasons (pp. 235–7).

Since the two views exist, confer intellectual satisfaction and cannot be rationally appraised, Simon concludes that the two views are equally necessary and equally justified.

It is no doubt the case that, considered as fundamental beliefs about the world, there is little room for rational debate between the reductionist position and its rival. But if we distinguish between metaphysics and heuristics (a distinction which Simon fails to make), surely there is room for rational appraisal of the relative merits of the two positions? The reductionist heuristic suggests that scientists look for explanations of biological phenomena in physico-chemical terms. Research on these reductionist lines has, as is well-known, been forging ahead, while that on anti-reductionist lines is, and has been, stagnating.

Simon agrees that the reductionist approach has proved fruitful, especially in the field of genetics. He even endorses the tentative claim of genetics to complete reducibility:

> There is no known biological fact concerning the transmission of heredity for which there is not at least strong evidence of its explicability in terms of chemistry (p. 145).

Simon admits (p. 145) that, at least in genetics, where explanations may be given ultimately in terms of chemistry, when biological statements are used, they are used purely for *convenience* in certain situations. That is, they contribute nothing to explanation which is not already covered by the explanation in terms of chemistry; once an adequate explanation in terms of chemistry has been given, the biological statements are redundant. But if this is the case, what are we to make of Simon's complementarity argument? If the anti-reductionist approach is not necessary to genetics, in the sense of describing another aspect of the phenomena under investigation, why should it be irreplaceable in other branches of biology?

There is a paucity of literature in this field and, in fact, no good introduction to the philosophy of biology exists. This book, written in a clear and simple style by someone who obviously has a sound knowledge of biology, touches on all of the major problems in the philosophy of biology. It is only unfortunate that the book is not grounded on a unified, critically defended, general philosophy of science.

JENNIFER GARNER
London School of Economics

SKLAIR, L. [1973]: *Organised Knowledge: A Sociological View of Science and Technology*. London: Hart-Davis, MacGibbon. £2.95. Pp. 284.

The basic point which Mr Sklair hopes to make is simply this: Science-technology is an important *institution* in our society. Yet philosophers of science (with the possible exception of Kuhn) have totally neglected the new epistemological dimensions which arise when knowledge-seeking becomes organised—and funded. Even sociologists (Merton is singled out for special criticism) have tended to view science as an enterprise which is so special that it should not be submitted to ordinary sorts of institutional analysis.

Sklair feels that mistaken conceptions about science are of more than academic importance. One of the greatest problems facing both highly industrialised and technologically underdeveloped countries today is that of how to subject science-technology to democratic scrutiny and direction. By perpetrating the myth that the scientific enterprise is autonomous, disinterested and independent of its social environment, by uncritically assuming that any increase of knowledge is an absolute good, and by reinforcing the scientists' claims that only they can make informed decisions about science, philosophers and sociologists have unwittingly helped construct the scientific-technological juggernaut of modern times. Sklair calls for the construction of a *social* philosophy of science along neo-Marxian lines.

These are heady issues and Sklair spends more time taking pot-shots at scholars and throwing out suggestions for future research than in articulating a position of his own. Since there are resumés of a wide variety of empirical studies as well as attacks on the orthodox theoretical views, this book provides a readable and provocative introduction to much current work in sociology of science.

Much of the experimental research seems to consist in the collecting of quantitative data which corroborate common-sense theses. Thus we learn that indeed there are more women students in pure science than in engineering (9·5 per cent *vs.* 0·5 per cent in the study cited) and that the *Times* devotes more column inches to science-related news that does the *Express*, *Mirror* or *Standard*. (Figures for the *Guardian* and *Telegraph* were not included for some reason.)

As far as pure sociological theory is concerned, most of the exact quantitative details are unimportant and one might be inclined to view this as rather useless data collecting. Yet the accumulation of these figures can have a considerable impact on our political awareness. Only 3 per cent of the news in the *Times* is even vaguely related to science. The U.S.A. spent 16 per cent of its 1965 federal budget on *R* and *D*. Ninety per cent of the scientists today live in fourteen countries. In order to get patent rights for the production of drugs such as tetracycline, Columbia had to buy intermediate products from the pharmaceutical firms at prices 82·6 per cent above the world market rate. Concrete details such as these provide evidence for the general impression about the role of science in the modern world which Sklair is trying to make.

Sklair's most detailed criticism is directed against the Mertonian theory of the norms of science. According to Merton, modern scientific inquiry has, since its Puritan beginnings, been dominated by a certain ethos. Upon analysis we find four basic norms: universalism (no aspect of science, including choice of research problem, should be ethnocentric), communism (scientific results belong to everyone), disinterestedness, and organised scepticism.

Merton goes on to claim that these norms are essential for the functioning of science; therefore any society which interferes with these aspects of scientific inquiry will seriously inhibit the progress of science. Merton also believed that the ethos of science was a manifestation of the highest standards of civilised man and that only in a democratic society could the ethos of science (and therefore science itself) survive.

Sklair takes issue with each of Merton's claims. Attitudinal studies show that many scientists do not in fact subscribe to Mertonian norms. Furthermore, there is no correlation between the scientific productivity of a scientist and the degree to which his or her value system is Mertonian. Scientists in industry neither choose their own problems nor make their results public. Kuhnian normal scientists do not exemplify the norms of impartiality and scepticism.

Sklair concludes that the Mertonian norms are not functional imperatives for science. Neither are they as morally laudable as Merton believes. For the norm of universalism is 'held to exclude any extra-scientific criteria in the prosecution and evaluation of scientific work' (p. 111). Of course Sklair thinks that social considerations *should* enter in—especially when the scientist is choosing an area of research.

At first one may feel that Sklair has romanticised and exaggerated Merton's position. However, this doubt is quickly dispelled upon reading 'Science and the Social Order', a lecture which Merton gave in 1937 on the state of science in Germany. This essay, which has been reprinted and also appears as part of one of Merton's books, comes close to saying that it is the norms of pure science which pose the biggest threat to a totalitarian state. However, it might be argued that it was through abandoning most of the Mertonian norms and working in the Manhattan project that scientists produced the most efficacious deterrent to Fascism!

The reader may well feel that it is high time one drew the distinction between pure and applied science. Surely Merton's norms were only intended to apply to basic research. And are not Sklair's concerns about democratic scrutiny really more appropriate for applied science?

However, Sklair does not find this distinction of much use. For science policy questions, the most important distinction is between expensive science and cheap science. Sociologically, the place where a scientist works (*e.g.* industry *vs.* university) may be more important than the type of research done. And one cannot understand the historical development of pure science without also studying technological capabilities.

By choice of title as well as by its claims, Sklair's *Organised Knowledge* invites comparison with Popper's *Objective Knowledge* and Polanyi's *Personal Knowledge*. But I must report that the orthodox philosopher of science will find little of interest here. Sklair's twenty-page commentary on Popper, Lakatos, Kuhn and Polanyi is very weak. For example, he says that

> 'Lakatos, carrying the Popperian torch, but with his own new high-octane fuel, considers that science is a matter of permanent revolution . . . whereas Kuhn considers that . . . scientists are content to work away within a fairly settled context where the fundamentals are not questioned . . .' (p. 135).

Since the next paragraph footnotes *Criticism and the Growth of Knowledge*, we cannot explain this gaff in terms of ignorance of Lakatos's recent work. However, the philosopher who wishes to study the ethical and social dimensions of scientific enquiry may find the book of heuristic value.

NORETTA KOERTGE
Indiana University

BODEN, MARGARET [1972]: *Purposive Explanation in Psychology*, Harvard University Press. $13.50. Pp. 408.

This book is an imaginative and scholarly examination of questions concerning purpose, in descriptions of machines, brains and behaviour. The essential position to be defended is that mechanisms—including neurophysiological mechanisms—can only be purposive when they embody internal representations of the problem-space on which they act, and of their own potentialities for action. This approach may give a basis for cognitive psychology, and for machine intelligence. Such 'internal representations', or 'internal models' are left out of stimulus-response theories of psychology; but are here regarded as necessary for intelligence, and for accepting mind. Thus on page 3: 'Only if the concept of mind is interpreted in this sense can one understand how it is possible for the mind to act on the body'. The origins of this notion go back to Kenneth Craik's *The Nature of Explanation* (1943) and in important respects—which perhaps Miss Boden has shown for the first time—to William McDougall, whose views form a running theme and reference base for her book. Some background knowledge of the history of psychology is necessary to see the significance of the historical quotations, which, however, do give an extra level of interest and serve to relate the main issues to the ancient question of vitalism versus mechanism in considering the behaviour of organisms. This emerges (if the word can be used in this context) here most clearly in an extended discussion on instinct: again with special reference to McDougall's ideas, which are seen as germinal to modern ethology.

The concept of machines embodying representational models has come to life with the modern computer, and has been developed by several workers in machine intelligence, recently by Marvin Minsky whose contributions are described without unnecessary technical detail. There are also references to recent work on scene analysis by computer, inspired by Chomsky's generative grammar. This is at present an active technical research field having clear philosophical implications. Indeed, as the author shows, the technical success so far achieved has changed some of the (mainly negative) assertions which seemed self-evidently true to many philosophers (and specifically to McDougall with his vitalistic hormic force or fluid). This applies to much contemporary psychology, of the 'Third Force' variety such as some of the writings of R. D. Laing. The social point here is that although thinking about man in terms of machines is often regarded as dehumanising, and so degrading, this may no longer be true if we emphasise and appreciate the non-mechanistic qualities of symbolic internal representation, whether by machine or by the brain cells of organisms including man. Purpose and mind are attributed to the properties of *what* is represented, rather than to the characteristics (except limitations) of the representing mechanisms. This implies that psychological 'purpose' explanations of behaviour are not to be couched in terms of the mechanisms of neurophsiology, for *they* are not purposive—though the models they represent and act upon to guide behaviour may be purposive. This gives rise to several interesting philosophical puzzles, which though hot are handled cooly, with a clear head for the necessary detail and the broad issues which are highly significant to current psychology, machine intelligence and the philosophy of mind. Perhaps some of the issues are too hot, and have to be dropped. By implication consciousness (or awareness of situations), is not necessary for purpose: but if brains are 'merely' machines representing aspects of reality by their internal models, what should we say of consciousness? This question has often been asked, in various forms, but still waits for an answer.

R. L. GREGORY
University of Bristol

CORRECTION

We should like to point out the following mistakes in the list of 'Recent Publications', *this volume*, p. 92. The books *Philosophy of Physics* and *Method, Model and Matter* were written, rather than edited, by Professor M. Bunge.

INQUIRY An Interdisciplinary Journal of Philosophy and the Social Sciences
Edited by Arne Naess
and Alastair Hannay

Number 4 Winter 1974 Volume 17

Objectivity and Truth in History	J. L. Gorman
Beim Sternenlicht der Nichtexistierenden: Zür ideologiekritischen Interpretation des platonisierenden Antipsychologismus	J. L. Nyiri
Freud and the 'Homeric' Mind	Jerry S. Clegg
Discussions:	
Correct vs. 'Merely True' Act Descriptions	Arthur R. Miller
On Miller's Paradoxes and Circles	David Rayfield
Rorty on Belief and Self-Deception	Béla Szabados
Review Discussion:	
Popper's *Objective Knowledge*	Paul Feyerabend

Published quarterly by: Universitetsforlaget, P.O. Box 307
Blindern, Oslo 3, Norway

Annual Subscription: Institutions – N. kr. 75.00, £4.90, $12.50
Individuals – N. kr. 60.00, £3.90, $10.00
Postage included

Philosophy and Phenomenological Research
AN INTERNATIONAL QUARTERLY
Edited by MARVIN FARBER

SEPTEMBER 1974, VOLUME XXXV, No. 1

CONTENTS

I. ARTICLES
  - Totalities and the Logic of First Cause Arguments, by CRAIG HARRISON
  - A Common Misunderstanding Concerning Husserl's *Crisis* Text, by PHILIP J. BOSSERT
  - The Concept of Action in the Bhagvad-Gita, by D. C. MATHUR
  - Armstrong and Strawson on 'Disembodied Existence', by MAX HOCUTT
  - The Relation Between Knowing and Making as an Epistemological Principle, by K. KUYPERS
  - On Reflection and Negation, by R. A. MALL

II. DISCUSSIONS, REVIEWS

III. COMMUNICATION

IV. NOTES AND NEWS, RECENT PUBLICATIONS

Subscription prices $8.00 per year for individuals; $10.00 per year for institutions. Separate issues, $2.00 and $2.50 respectively. Postage and handling for back issues only. Papers submitted for publication will not be returned unless accompanied by a self-addressed stamped envelope, or equivalent postage.

A Circular listing the main contents of the Journal since 1940 will be sent upon request.

Address all communications to

PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH
State University of New York at Buffalo, Buffalo, New York, 14226

Noûs

Nihil philosophici a nobis alienum putamus

A Quarterly Journal of Philosophy

edited by Hector-Neri Castañeda

Philosophy Department, Indiana University

Bloomington, Indiana 47401

NOÛS publishes quality essays and brief discussions on philosophical problems regardless of the author's philosophical school or point of view. From time to time, NOÛS will publish symposia, surveys of work recently done on specially selected philosophical topics, critical studies of recent books (rather than reviews), and catalogues of recent publications in philosophy. NOÛS publishes the symposium papers of the A.P.A. Western Division meetings in its March issue.

A quarterly, NOÛS is published in March, May, September, and November each year. For *individuals* the subscription rates are: one year, $9.00; two years, $17.00; three years, $25.00. For *bona fide students:* $8.00 per year. For *institutions:* $16.00 per year (plus 10% for subscriptions handled through a subscription agency). Add 50 cents per year for all non-U.S. subscriptions. Make checks for one year subscriptions payable to INDIANA UNIVERSITY NOÛS. Checks for multiple year subscriptions should be made payable to THE INDIANA UNIVERSITY FOUNDATION-NOÛS.

VOLUME VIII, NUMBER 3, SEPTEMBER 1974

CONTENTS

ROBERT MERRIHEW ADAMS—*Theories of Actuality*

SAMUEL C. WHEELER III—*Inference and the Logical 'Ought'*

JOHN W. SNAPPER—*Contextual Definition: What Frege Might Have Meant But Probably Didn't*

C. J. F. WILLIAMS—*Believing in God and Knowing that God Exists*

ROBERT P. MCARTHUR—*Factuality and Modality in the Future Tense*

MICHAEL D. RESNIK—*A Note on Interpreting Theories*

HOWARD BURDICK—*On the Problems of Abstraction and Concretion*

Announcement

Recent Publications in Philosophy

Order from

The Secretary, NOÛS

Department of Philosophy

Indiana University

Bloomington, Indiana 47401

RATIO

edited by Stephan Körner

VOLUME XVI DECEMBER 1974 No. 2

RUDOLF HALLER	Concerning the So-called 'Münchenhausen Trilemma'
MORTON WHITE	On the Use of English 'ble'-words to Express Possibility and Necessity
GORDON BAKER	Criteria: A New Foundation for Semantics
M. W. BEAL	Essentialism and Closed Concepts
HILLEL STEINER	The Concept of Justice
JOHN HARRIS	Leibniz and Locke on Innate Ideas
REVIEW:	
MICHAEL WELBOURNE	*The Nature of Things* (Anthony Quinton)

Price £2.25 net **Annual Subscription: £4.00 net**

Orders to: Journals Department, 108 Cowley Road, Oxford OX4 1JF, U.K.

Published by

BASIL BLACKWELL · OXFORD

Revue Internationale de Philosophie

Direction Scientifique: Professeurs PH. DEVAUX, de l'Académie royale de Belgique
H. GOUHIER, de l'Institut de France
CH. PERELMAN, de l'Académie royale de Belgique

Direction Administrative, Rédaction: MAX SERVAIS
99 avenue de l'Université, 1050 Bruxelles
Tél. 02/648 03 70

LA REVUE INTERNATIONALE DE PHILOSOPHIE consacre, depuis 1938, chacun de ses 4 numéros annuels à l'étude d'un philosophe ou d'un mouvement philosophique.

No. 107–108—1974—Fascicule 1–2

LA PHILOSOPHIE JAPONAISE CONTEMPORAINE

Sommaire

M. TANAKA, *Future*
S. OHE, *Japan in a World-Historical Perspective*
M. YAMAZAKI and S. SAITO, *Modern Japan in its Philosophical Development*
Y. TAKEUCHI, *Marxism in Japan*
M. MATSUMOTO, *The Absolute, Relative and Nothingness*
S. OHE, *Wert und Menschliches Sein*
T. IWASAKI, *A Criticism of Phenomenalism*
T. IMAMICHI, *L'humanisme, l'expression et l'idée du beau*
T. FURUKAWA, *A Philosophical Heritage of Japan*
T. IZUTSU, *The Philosophical Problem of Articulation in Zen Buddhism*
H. NAKAMURA, *Methods and Significance of comparative Philosophy*
VARIETE—ANALYSES ET COMPTES RENDUS—OUVRAGES RECUS—REVUE DES REVUES

Derniers numéros parus: 103-Etudes philosophiques et Informatique 104-165-Meinong 106-Lukacs

Numéros prévus: Problèmes actuels de l'Esthétique-Cassirer-Méthode et Philosophie de l'Histoire-Grabman

Abonnements:	Belgique:	530 F.B.
	Pays européens:	580 F.B.
	Autres pays:	620 F.B.

Philosophy of Science

OFFICIAL JOURNAL OF THE PHILOSOPHY OF SCIENCE ASSOCIATION

Editorial Board

Richard S. Rudner (Editor-in-Chief), Peter Achinstein, Gustav Bergmann, May Brodbeck, Arthur Burks, C. West Churchman, Theodosius Dobzhansky, Adolf Grünbaum, Edward Madden, Henry Margenau, Gerald J. Massey, Alberto Pasquinelli, Wesley C. Salmon, Kenneth F. Schaffner, Dudley Shapere, J. J. C. Smart, Melford E. Spiro, Patrick C. Suppes, Lewis K. Zerby, Peter D. Asquith (Managing Editor).
Book Reviews, James Van Evra.
Assistant Editors, Patricia Ann Fleming and Paul C. L. Tang.

Vol. 41 June, 1974 No. 2

CONTENTS*

The rationality of scientific discovery, part I: the traditional rationality problem, NICHOLAS MAXWELL / *Progress as a demarcation criterion for the sciences*, PAUL M. QUAY, S.J. / *Maxwell's demon and computation*, RICHARD LAING / *Discussion: Kuhn, Scheffler and objectivity in science*, JACK W. MEILAND / *Discussion: The quartercentary model of D-N explanation*, DALE A. THORPE / *Discussion: Inductive immodesty and lawlikeness*, JUHANI PIETARINEN / *Discussion: Van Fraassen's modal model of quantum mechanics*, NANCY CARTWRIGHT / *Discussion: Polanyi on the Meno paradox*, MICHAEL BRADIE / *Discussion: On a problem in conditional probability*, A. I. DALE / *Announcement* / *Book Reviews*.

* The contents of *Philosophy of Science* are indexed in *Science Citation Index*, *ABC Pol Sci*, *Language and Language Behavior Abstracts*, and *The Philosopher's Index*.

PHILOSOPHY OF SCIENCE is published quarterly by the Philosophy of Science Association, 18 Morrill Hall, Department of Philosophy, Michigan State University, East Lansing, Michigan 48824. Subscription price £17.50 (£18.50 non-U.S.A.) per Volume. Members of the Association receive the Journal as a privilege of membership paid for out of membership dues. Inquiries about subscriptions or membership should be directed to the Managing Editor.

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by

NICHOLAS RESCHER

VOLUME 11/NUMBER 4 OCTOBER 1974

CONTENTS

- I EDWIN CURLEY: *Recent Work on 17th Century Continental Philosophy*
- II PETER ACHINSTEIN: *The Identity of Properties*
- III ROBERT SOLOMON: *Hegel's Epistemology*
- IV RICHARD SCHACHT: *On 'Existentialism', Existenz-Philosophy and Philosophical Anthropology*
- V SOUREN TEGHRARIAN: *Linguistic Rules and Semantic Interpretation*
- VI VIRGIL C. ALDRICH: *Sight and Light*
- VII ROBERT HOFFMAN: *Vulgarity*
- VIII JOHN BEVERSLUIS: *Socratic Definition*
- IX DALE GOTTLIEB: *Foundations of Logical Theory*

Annual Subscription: $18.00 to Institutions (*ex*cluding Annual Monograph, price $7.00)
$8.00 to Individuals (*in*cluding Annual Monograph)

Published by BASIL BLACKWELL

Orders to: 108 Cowley Road, Oxford OX4 1JF, England

28 FEB 1975

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1974-5

President: Professor J. W. N. Watkins
President-Elect: Professor D. J. O'Connor
Hon. Secretary: Helena Sheiham
Hon. Treasurer: Dr D. A. Gillies
Members: Dr M. Machover, Mr H. Martins, Mr N. Maxwell, Dr E. Zahar

EDITORIAL BOARD OF THE JOURNAL

Professor R. B. Braithwaite
Professor H. Dingle
Professor Sir John Eccles
Professor Sir Harold Jeffreys
Professor W. C. Kneale
Professor S. Körner
Professor M. Polanyi
Professor Sir Karl Popper
Professor G. J. Whitrow
Professor J. H. Woodger

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules :

' The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all persons who have been proposed by two members, approved by the Committee, and elected at an ordinary meeting. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £1·00 ($2.50). Those who wish to join should write to the Hon. Secretary, Helena Sheiham, Department of Philosophy, London School of Economics, Houghton Street, London, WC2A 2AE.

Brit. J. Phil. Sci. Vol. 25 No. 3 September 1974

The British Journal for the Philosophy of Science

28 FEB 1975

Contents

Articles

209–221 George Schlesinger The Unpredictability of Free Choices.

223–234 John F. Post Propositions, Possible Languages and the Liar's Revenge.

235–259 Peter Urbach Progress and Degeneration in the 'IQ Debate' (II).

Discussions

261–265 Stephen Spielman On the Infirmities of Gillies's Rule.

265–271 M. L. G. Redhead On Neyman's Paradox and the Theory of Statistical Tests.

271–284 Fritz Machlup Situational Determinism in Economics.

285–288 A. W. Coats Situational Determinism in Economics: The Implications of Lastis's Argument for the Historian of Economics.

289 I. J. Good A Correction Concerning Complexity.

Reviews

291–300 David Miller on N. Rescher: *The Coherence Theory of Truth*; Jennifer Garner on M. A. Simon: *The Matter of Life*; Noretta Koertge on L. Sklair: *Organised Knowledge*; Richard Gregory on M. Boden: *Purposive Explanation in Psychology*.

Printed in Great Britain at the University Press, Aberdeen

Volume 25 Number 4 December 1974

BJPIA5 25 (No. 4) 301-402 (1974) 0007-0882

The British Journal for the Philosophy of Science

CALCUTTA * UNIVE

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Editors: JOHN WATKINS and JOHN WORRALL

Assistant Editors: COLIN HOWSON and ELIE ZAHAR

Subscriptions

The *Journal* is published quarterly. The subscription price is £7.00 net (U.S.A. and Canada U.S. $22.00) per volume post free; single issues £2.50 net (U.S.A. and Canada U.S. $8.00) plus postage. The subscription is £4.50 (U.S.A. and Canada U.S. $14.00) for members of the British Society for the Philosophy of Science, who should order direct from the publisher. (Membership of the Society is international and is by no means confined to people resident in the United Kingdom.) Orders may be sent to any bookseller or direct to Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT.

Notes to Contributors

Articles, correspondence and books for review should be sent to the Editor, John Worrall at the Department of Philosophy, London School of Economics, Houghton Street, Aldwych, London WC2A 2AE. Articles should be accompanied by abstracts of about 100 words.

Material is accepted for publication in the *Journal* on the condition that copyright is assigned to the British Society for the Philosophy of Science.

Three copies of all contributions should be sent, typed on one side only with wide margins and double-spaced. Footnotes, which should be as few as possible, should be typed separately, double-spaced at the end of the text. They should be referred to by consecutive numbering throughout the text, in which they will be inserted in page proof. Bibliographical references should be listed alphabetically at the end. The style of references should follow current *Journal* practice.

Long quotations are printed in small type without quotes, and should be so marked in the typescript. Single quotes should be used to form the names of words and sentences, and double quotes for all other purposes. Diagrams, tables, and illustrations should be on separate sheets, with their desired position in the text clearly indicated. Symbolic formulae appearing as separate lines should be clearly distinguished in the text, and foreign letters (Greek, German, etc.) indicated in the margin. Material to be printed in italics, especially English letters used as symbols ('*A*', '*x*', etc.), should be underlined. On matters of style and layout, authors should consult CHAUNDY, T. W. *et al.* [1964]: *The Printing of Mathematics*, London: Oxford University Press.

Offprints of articles may be ordered when corrected galley proofs are returned. Fifty are supplied free ; the price of extra copies is given on the order form.

© British Society for the Philosophy of Science 1974

For permission to reproduce material from *The British Journal for the Philosophy of Science*, please apply to the Editors.

Back Volumes

Volumes 1-16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17-24 may be obtained from Cambridge University Press, Bentley House, 200 Euston Road, London NW1 2DB.

Brit. J. Phil. Sci. **25** (1974), 301–318 *Printed in Great Britain*

Uniformity and Induction

by JOHN C. GRAVES

In Book III, Chapter III, Section 1 of his *System of Logic*, John Stuart Mill puts forward the principle of the uniformity of nature as a supposed justification of induction. Mill's principle has been cited in most subsequent discussions of induction and its justification. However, it is usually promptly dismissed as being based on a logical fallacy, or as hopelessly vague, or as unnecessary and incapable of doing the job it is supposed to do. I propose to argue, on the contrary, (*1*) that given Mill's very explicit statement of what the principle is and is not supposed to do, and his interpretation of the significance of deduction in science, there is no fallacy; (*2*) that the principle (in some form) is needed to fill a gap in the arguments used in most other discussions of induction; and (*3*) that while Mill's own formulation is vague, it is not hopelessly so, and at least in special cases it can be made quite precise by using some of Carnap's ideas. In particular, I want to defend Mill from the criticisms made by Nagel, Strawson, and Reichenbach.

Mill begins by asserting that the principle is implied in the very statement of what induction is: it is an assumption about the course of nature and the order of the universe, *i.e.* an ontological assumption, and in content it asserts that there are parallel cases in nature, or that what happens will be repeated whenever there are sufficiently similar circumstances.[1] He then goes on to say that this principle, which may be put simply as 'Nature is uniform', is both borne out by experience and *warrants all inferences* from experience.[2] It is Mill's use of the term 'warrant' that leads one to believe that he is offering uniformity as a wholesale justification of induction, especially since he also describes it as the 'fundamental principle, or general axiom of Induction'[3].

But Mill immediately introduces an important qualification. He asserts[4]

> It would yet be a great error to offer this large generalization as any *explanation* of the inductive process. On the contrary, I hold it to be an instance of induction, and induction of by no means the most obvious kind. Far from being the first induction we make, it is one of the last, or at all events one of those which are latest in attaining strict philosophical accuracy. As a general maxim indeed, it has scarcely entered into the minds of any but philosophers; nor even by

Received 7 August 1973

---

[1] Mill [1898], pp. 200–1.

[2] *Ibid.*, p. 201.

[3] *Ibid.*, p. 201.

[4] *Ibid.*, p. 201.

> them, as we shall have many opportunities of remarking, have its extent and limits been always very justly conceived. The truth is, that this great generalization is itself founded on prior generalizations. (My italics.)

And Mill realises that this creates a problem, namely, in just what sense can such a clearly empirical, inductively derived principle be a warrant for any other inductions? If one were to interpret 'warrant' as providing a premise from which particular inductive conclusions could be *proved* by valid deductive inferences, the overall argument would certainly be circular. Since Mill does go on to refer to uniformity as the 'ultimate major premise of all inductions', such commentators as Madden have concluded that Mill was indeed guilty of this fallacy.[1]

But despite the somewhat misleading title of his book, Mill was not trying to develop a system of logic in the modern sense of a formal system of axioms and rules of inferences for proving theorems deductively. He sought rather to provide an 'organon', in the same sense as Aristotle or Francis Bacon—an instrument for effective reasoning in the sciences, and indeed, on moral questions. Thus he does not add to the formal theory of the syllogism, but asks instead what role syllogistic reasoning might play in science.[2] He concludes that *any* inductive argument can be thrown into a deductive form, albeit somewhat artificially. However, by itself this move contributes very little, since the general statement that it uses as major premise is known to be true only for those particular cases that provide its evidence, and which are included in the minor premise. Any inference, therefore, really moves from particulars to other particulars.[3] Syllogistic or deductive reasoning does not prove conclusions. However, it reveals the general form or mode of reasoning used in moving between the sets of particulars, and it prevents one from drawing conclusions based on the peculiar features of any of the particulars, rather than the common features which link them together as a class.[4] Finally, it shows what would have to be *assumed*, whether or not the assumption is correct: 'the major premise of a syllogism stands to the conclusion [by] not contributing at all to prove it, but being a necessary condition of its being proved'.[5]

If we meet Mill on his own terms we see that there is no more circularity here than in any other deductive argument used in science. For he claims that in all such cases the major premise and conclusion are both inferred equally from the evidence in the minor premise.[6] It is *not* the possibility of using a deductive form that justifies induction, and the use of such forms gives *no* additional certainty. Nagel is entirely wrong in claiming that Mill supposed 'that in some cases at least the sciences do achieve complete and final certainty concerning the laws of nature', or that he was less subtle and consistent than Hume.[1] While Hume did wish to show the impossibility of reaching absolute certainty about matters of fact, Mill realised that theoretical scepticism was useless for science. Since some statements must be accepted in practice anyway, Mill was concerned instead with the 'moral certainty' that Hume himself had allowed. We know that it is always possible that our inductive predictions may fail, and thus any putative justification that ruled this out would have to be incorrect, but we can not worry about this constantly. Likewise Nagel's claim that Mill was inconsistent with his own atomistic empiricism does not hold. It is not inconsistent but in fact ***required*** by his empiricism that any principle used in science must itself be justified inductively.

---

[1] Madden [1960], p. 289; also Nagel [1950], p. xlii.

[2] Mill [1898], Book II, Chapter III, pp. 119–36.

[3] *Ibid.*, p. 122.

[4] *Ibid.*, p. 129.

[5] *Ibid.*, p. 202.

[6] *Ibid.*, p. 374.

According to Mill, what actually happens seems to be this: when mankind made its first inductions, there was neither the possibility of nor the recognised need for any justification. People simply made them, whether from natural instinct or Humean habit.[2] Gradually, we became aware of the fact that inductive predictions were remarkably successful in a wide variety of cases, and that our inductions could be organised in a hierarchy of increasing generality. Uniformity of nature was finally seen as the apex of his hierarchy. Suppose we observe an additional black crow, having seen black but never having seen any non-black crows. This confirms the hypothesis that all crows are black. But it also confirms that all crows have the same colour, that all crows have some properties in common, that there are such things as crows or fixed species in nature generally, and so on up to the claim that nature does exhibit considerable uniformities. We became aware of the truth of uniformity through its being simultaneously confirmed along with the lower-level hypothesis on which our immediate attention was centred. For Mill, scientific induction must be founded on spontaneous induction, and an appeal to uniformity just reflects our self-conscious awareness of the full implications of our methodology.[3]

In order to evaluate the charges of vagueness, uselessness, and logical inadequacy in more detail, let us consider just where and how Mill uses the principle of uniformity. Too often writers who discuss induction have quite different forms of reasoning in mind, and their justifications are appropriate only for that particular form. Mill identifies induction with generalisation from experience, and specifically with the following sort of case: we have observed several members of a class, and found that they all have some property in common, with no exceptions. We then wish to infer that all members of that class have the property in question. We may express the minor premise, our evidence, in the form $Pa_1, Pa_2, \ldots, Pa_n$, where P is some definite property, and the various a's are names of individuals. The value of the total number n is not relevant here. Our conclusion then takes the form $(x)\, Px$, and our problem of logic is whether there is a true major premise which can be combined with the minor premise to give this conclusion deductively. In view of his insistence that all inference is from particulars to other particulars, and his pragmatic emphasis on successful predictions, it would seem that Mill is really interested in the next case, Pa_{n+1}, as his conclusion, rather than $(x)\, Px$. However, for Mill the very meaning of induction requires that the argument be equally good for any other particular Pa_{n+m}.[1] This will work only if the conclusion has universal form. A statistical conclusion does not allow us to make categorical predictions about any particular in its reference class, and thus lies outside of Mill's field of interest.

---

[1] Nagel [1950], p. xlii.

[2] Mill [1898], p. 204.

[3] *Ibid.*, p. 208.

In Mill's own example, the minor premise is 'John, Peter, *etc.*, are mortal', and the conclusion is 'All men are mortal'. As a possible major premise, Mill proposes 'What is true of John, Peter, *etc.*, is true of all mankind'.[2] The vagueness of the expression 'what is true . . .' is best symbolised by a free predicate variable F, giving $Fa_1, \ldots, Fa_n \,.\! \rightarrow (x)\, Fx$. However, this is just a statement form, rather that a statement, and nothing can yet be inferred. At this point the logician knows only three moves. (*1*) He can quantify universally, giving $(F)\,(Fa_1, \ldots, Fa_n \,.\! \rightarrow (x)Fx)$, *i.e.* 'everything that is true of John, Peter, *etc.* is true of all men'. Using this we can validly deduce $(x)\, Px$; however, the major premise is of course false. (*2*) He can quantify existentially, giving $(EF)\,(Fa_1, \ldots, Fa_n \,.\! \rightarrow (x)\, Fx)$, *i.e.* 'something that is true of John, Peter, *etc.* is true of all men'. Here the major premise is certainly true: if necessary, we can use the defining property of the reference class. However, the conclusion now cannot be inferred. (*3*) He can instantiate F as P, giving $Pa_1 \ldots Pa_n \,.\! \rightarrow (x)\, Px$. Here again the deduction is valid, but we do not know whether the major premise is true or false, and in any case it begs the question. For this reason it has usually been concluded that the search for a demonstrative syllogism is hopeless.

I do not think Mill would accept this analysis or interpretation. He certainly knows very well that many inductions do prove false, although a remarkable number succeed. In fact, he takes great pains to emphasise the diversity and complexity of nature as well as its uniformity, and criticises Bacon for his more naive use of induction by simple enumeration.[3] Furthermore, it is much too crude to say that nature either is or is not uniform, as if we had a simple dichotomy. What we have is a great many related particular uniformities, and the overall uniformity of nature is nothing more than the 'tissue' of these partial uniformities, or laws.[1] We know from past experience that some uniformities do exist, and whenever we do argue inductively we are assuming (or believing, or hoping) that the case in question is one of them. We may of course turn out to be wrong; but to assume beforehand that we will be wrong is indeed inconsistent with the principle of uniformity.

---

[1] Mill [1898], p. 186. [2] *Ibid.*, p. 202. [3] *Ibid.*, pp. 203–4.

Let us now try to summarise Mill's claims about uniformity:

A. It is that principle which would have to be true if an inductive syllogism were to be valid, *i.e.* it is a necessary presupposition of its validity. This form makes it look like the end product of a Kantian transcendental argument. However, we must note that the argument is in the subjunctive mood. We do not yet know either that the argument is valid in the case of interest or that the principle is true. But if it were true, it would have this effect.

B. Pragmatically, in the actual practice of scientists, it is always *taken* to be true, at least implicitly, whenever any conclusion is inductively inferred. Once again, this might be the wrong thing to do, and mere fact that it is so used does not *ipso facto* justify it. But in so far as one makes any predictions based on past experience, it is uniformity rather than some alternative principle that determines the actual content of these predictions.

C. It has in fact been found to be true in the past. We have discovered many previous uniformities empirically and inductively. In some cases we may already have verified certain uniformities by complete enumeration, if in fact there were only a limited number of observable cases. And even if our inductive prediction fails on the next try, the fact that we have previously observed many cases, all of which were similar rather than having randomly distributed characteristics, *already* establishes a significant degree of uniformity.

One might accept that Mill's three claims are all correct, but still ask how they add up to a justification. Mill recognises that any justification for a principle used in science is not logical, but ontological and pragmatic. We must not look up a chain of deductive premises to some supposed self-evident or *a priori* truth, though we may find it convenient to exhibit the form of our reasonings in this way. Rather we must turn to the facts about the world as a whole and the need of scientists for decisions and actions. *Certainty* is not an essential part of justification but *success* in predicting and dealing with the real world is. If we ask, 'What must the world be like if induction is ever to work?', the answer is that it must have some uniformity. If we ask, 'What must the world be like if induction is to work most of the time, or in some specified percentage of cases?', the answer is that it must have some particular, and fairly high degree of uniformity. And our past experience shows it to have this high degree, so we in fact assume it in making inductive predictions confidently.

---

[1] Mill [1898], pp. 206–8.

Mention of 'degree of uniformity' raises the problem of what such degrees might mean, or whether we could determine this degree, and perhaps use it to decide how confident we should be in making predictions. Certainly there is no intuitive notion of how one might establish a scale, but still most people would agree on comparative judgments. A world in which all crows are black is more uniform than a world in which all but one are black, which is more uniform than one in which is some comparable ration of black and white crows, which is more uniform than one in which crows are randomly distributed in all the colours of the rainbow, which is more uniform than one in which crows have no common properties, or there are no fixed species, and so on. In turn it is less uniform than one in which crows have all their properties in common except for some single individuating feature, which is less than one in which all things are crows, which is less than one in which everything is part of an undifferentiated Parmenidean whole—perhaps the extreme limit of complete uniformity. Even if a more precise analysis might show this uniformity to be multidimensional, the principle is not so vague as to prevent us from speaking of more or less uniform in comparing conceivable situations.

At this point Nelson Goodman's criticism may raise its ugly head. One might inquire whether a world in which all emeralds are grue is just as uniform as one in which all emeralds are green. If this were true, one might conclude that the notion of uniformity is vacuous, since no matter how complex and apparently non-uniform the actual world seemed to be, one could always (at least after the fact) define some artificial set of grue-like predicates such that the world would have an arbitrarily high degree of uniformity with respect to these predicates.

I do not propose to solve Goodman's paradox here, though I believe its resolution is important for any notion of uniformity. Mill himself would undoubtedly have resolved it through his belief in the existence of natural kinds, which were not mere distinctions for convenience but radical distinctions in the things themselves, and thus provided the ontological basis for 'uniformities of coexistence'.[1] Nevertheless, Mill must surely recognise that in the history of science we have often changed our opinions of what these natural kinds were, and changed the concepts in which the laws or uniformities of nature were expressed. As the conventionalists have emphasised, it is quite possible, and it sometimes happens, that when an induction fails we do not conclude that nature has no uniformity in this respect, but look instead for more subtle concepts and complex laws.

---

[1] Mill [1898], p. 379 and Book I, Chapter VII.

However, as Goodman's own resolution in terms of entrenchment and actual projections suggests, this procedure occurs only very rarely, in comparison with the overwhelming number of times when induction with respect to entrenched and relatively simple predicates succeeds.[1] This is a contingent but vitally important fact about our world, and the possibility of our gaining knowledge of it. If we had to revise our scientific concepts continually to preserve uniformity the notion would still be logically possible but practically useless, and in any case it could only be used after the fact. Quine has made the very intriguing suggestion that uniformity and simplicity are related notions, and that both may be closely related to the limitations of and biases inherent in our perceptual apparatus and experimental techniques. We simply do not perceive something as a uniformity unless it appears simple to our human understanding, even if it might be admissible from some other viewpoint.[2] 'Grue' involves changing our perceptual criteria in a way that 'green' does not, even if they are logically symmetrical. While the growth of science allows us to be increasingly sophisticated about what can count as a uniformity, and we do adjust our criteria to make them fit experience better, we never go terribly far. Mill's claim may be thus strengthened: Nature is uniform not just in some abstract sense, but with respect to relatively simple human concepts. Alternatively, we can say that the 'phenomenal world', the world that we encounter experientially, is uniform, and these are certainly contingent truths.

It should be noted that uniformity is certainly not a panacea for solving all the problems about induction. Cohen and Nagel criticise Mill on the ground that it is useless to appeal to the principle in any particular investigation, since it does not specify in which particular ways nature is uniform.[3] But Mill is only too aware of this. He says explicitly[4]

> Why is a single instance, in some cases, sufficient for a complete induction, while in others, myriads of concurring instances, without a single exception known or presumed, go such a very little way toward establishing a universal proposition? Whoever can answer this question knows more about the philosophy of logic than the wisest of the ancients, and has solved the problem of induction.

Mill never claims that he has reached such a general solution, and he certainly regards it as a crucial task for further investigation in inductive logic. But he wants to separate the *ontological* question, What must the world be like if induction is to work at all? from the *epistemological* question, When should we decide to accept a particular inductive argument in terms of the number and variety of observed instances? Uniformity is an answer, and indeed the *correct* answer to the former, but it is quite irrelevant to the letter. But both questions are important, and if the former is easier, it still must be settled before we can deal with the latter. Much of the rest of the book is devoted to trying to develop more detailed rules for such an ideal inductive logic. For example, Mill emphasises such factors as causal connectibility and prior plausibility in choosing which inductions to accept.[1] Nevertheless, even these must be founded on previous experience and discovered inductively. We can test perceptual experience from one of our human faculties, but only against that from another. There is no external court of appeal beyond experience and induction.

---

[1] Goodman [1955], Chapter V, pp. 87–120.

[2] Quine [1963].

[3] Cohen and Nagel [1934], p. 268.

[4] Mill [1898], p. 206.

In direct contrast to Mill, Strawson has argued that induction neither needs nor can be given a general justification, whether from uniformity or any other single principle. One must note first that Mill and Strawson are not really talking about the same thing. In discussing induction Strawson refers to any kind of procedure in which we cannot prove a statement by a deductively conclusive argument, but where we weigh evidence as giving partial support to a hypothesis. These procedures may include a variety of specific rules, so Strawson ridicules the idea that it is meaningful to speak of 'The Inductive Method', or use it as the name of a particular mode of reasoning.[2] Since this covers all non-deductive cases, it is naturally implausible that some single principle should govern all of them. But Mill might readily accept this. His appeal to uniformity applies only to the specific case of reasoning from $Pa_1 \ldots Pa_n$ to $(x)\, Px$.

Nevertheless Strawson is willing to challenge Mill on his own ground. For he argues that Mill has confused or conflated what Strawson takes to be two quite distinct claims: (*I*) The universe is such that induction will continue to be successful, and (*II*) Induction is rational (reasonable).[3] Strawson's point is that (*I*) is an ontological claim, expressing a contingent fact about the world which might be discovered only empirically, while (*II*) is a necessary truth, derived from the meanings of the terms involved, and true in all possible worlds. Now Mill would undoubtedly accept (*I*) as an adequate expression of the fact of uniformity, but he would insist that (*II*) is equally contingent, that it is true if and only if (*I*) is true, and that (*I*) provides the justification of (*II*).

Strawson's mistake arises from his belief that justification is a logical matter, so that one can justify some set of rules only by showing that it follows from some other set. Thus if (the totality of) inductive rules were justified by appealing to themselves, the argument would be circular; if they were justified deductively, the argument would be unsound for the very reason that inductive rules are different from deductive in lacking the crucial feature of guaranteeing a true conclusion from true premises. He asks, rhetorically, Why is it that no one ever asks for a justification of deduction? The presumed answer is that such a justification would have to use deduction in a circular way, since deduction itself provides the very standard of justification, legitimacy, or rationality here. Strawson concludes that those who seek a justification of induction are simply unwilling to accept the force of Hume's argument, and want to sneak certainty in the back door again. One should recognise instead that in non-deductive cases, inductive rules themselves provide the only possible standards of legitimacy or rationality, and thus it is hopeless to try to transcend them.[1]

---

[1] Mill [1898], pp. 371, 209, *et passim.*

[2] Strawson [1952], p. 258.

[3] *Ibid.*, p. 261.

Strawson here misses the whole point of Mill's contention that justification is ontological and pragmatic, rather than logical. The request for a justification of deduction is absurd only if we think in terms of purely formal systems, where essentially arbitrary rules of inference tell us how to get from axioms to theorems. Insofar as such deductive systems are to be applied to the physical world, their application must be justified by facts about this world. It may be that for some particular deductive rule, we cannot conceive of an alternative and thus assume that it holds in all possible worlds. But *a fortiori*, it then holds in this world: a justification based on a fact about this world by no means rules out the possibility that it would also hold in other worlds in which similar facts obtained. Mill would say that deduction *in natural science* also needs a justification, and it is given the same one as induction, via facts discovered empirically. If one interprets this as an inductive justification of both induction and deduction, well and good. But the logical-ontological seems a more important dichotomy than the inductive-deductive here. For a concrete illustration of Mill's point, Reichenbach has proposed that facts discovered in the quantum realm can best be interpreted by replacing traditional two-valued deductive logic with a multivalued logic.[2]

Strawson's reason for separating (*I*) from (*II*) is his belief that (*II*) is purely linguistic. He says, 'The rationality of induction, unlike its "successfulness", is not a fact about the constitution of the world. It is a matter of what we *mean* by the word "rational" in its application to any procedure for forming opinions about what lies outside of our observations or that of available witness'.[3] And elsewhere, 'It is an *analytic* proposition, though not a proposition of mathematics, that other things being equal, the evidence for a generalisation is strong in proportion as the number of favourable instances, and the variety of circumstances in which they have been found, is great'.[1] But how does one discover these meanings, or determine that (*II*) is really analytic?

---

[1] Strawson [1952], pp. 248–50.

[2] Reichenbach [1944].

[3] Strawson [1952], pp. 261–2.

Such an identification of 'rational' with 'inductive' might well be true as a statement about the way in which 'rational' is actually used by scientists and others in non-deductive contexts. One might even find it in a dictionary, where the entries are empirical statements about normal usage. But this will not establish identity of meaning in the strong sense that Strawson needs. For 'inductive' is a purely descriptive term, a name given to a certain mode (or modes) of reasoning. 'Rational', on the other hand, is clearly an evaluative term. To call anything rational is to praise it, to recommend it as a way in which people *ought* to behave. If Strawson's argument depends on identity of meaning, it is simply an extremely blatant instance of the naturalistic fallacy. The meaning of 'rational' could not have been specified in advance by linguistic stipulation. If it is *now* used in the same way as 'inductive', this is only because in the long history of the race inductive methods have proved far more successful than any others in solving problems and making correct predictions. And this happened only because the world had a large degree of uniformity. Any analysis of linguistic usage must consider its historical evolution and the contexts in which it arose, not merely its present role.

Strawson's final argument against Mill's ontological justification is that we can conceive of a state of chaos in which all our ordinary inductive predictions would break down.[2] This is then a possible world, but one in which (*I*) is false. Mill also allows this possibility, as part of his argument for insisting that uniformity is not a necessary truth about the world.[3] However, Strawson contends that even here, (*II*) would be true. Induction would still be rational, for at least one could make the successful induction, when chaos has been observed in the past, expect (or predict) chaos in the future.

But this will not work either. Strawson deals only with one form of chaos. In an equally possible world, one which shifts from epochs of complete disorder to epochs of considerable order at random and unpredictable times, even Strawson's induction breaks down. Furthermore, in the world Strawson describes, expecting future chaos is not at all like expecting that the next crow will be black. It is purely negative, and not the basis of any positive predictions. Indeed, expecting chaos is pragmatically equivalent to not having any expectations at all. It would make no sense to talk of success, and it would make no sense to talk of rationality. Far from saying that induction is still rational in such a world, it would be better to say that *no* methods are rational. No one would use the term 'rational' unless he had some expectations of success.

---

[1] Strawson [1952], pp. 256–7.

[2] *Ibid.*, pp. 260–1.

[3] Mill [1898], p. 371.

Let us now consider how Mill's uniformity principle might help to supplement other approaches to induction. Reichenbach is primarily concerned with a different problem from Mill's, namely the problem of predicting the long-run limit of frequency of success in a sequence of trials, each of which may give success or failure. After any number of trials we will have a ratio of successes to total trials, and we must decide how to extrapolate from this observed ratio to the eventual limit (if a limit exists) which is then *defined* as the probability of success.[1] More precisely, he states, 'The aim of induction is to find series of events whose frequency of occurrence converges toward a limit'.[2] Given that problem, his proposal is that we use the 'straight rule' of taking the observed ratio, without any added terms, as a continually self-correcting estimate of the ultimate limit.

We see that Reichenbach is dealing with a more general question than Mill, since he is concerned with statistical sequences which will normally include both successes and failures, rather than unexceptioned universal laws. But it is incorrect to suggest, as Lenz does, that Mill's problem is just a special case of Reichenbach's.[3] This would be so only if we knew that the sequence of trials had some definite limit. If it could be extended indefinitely, it would be possible to have any finite number of failures in the early stages, so long as there were nothing but successes beyond some particular trial. In this case the limit ratio of successes to trials would approach 1, but it would be false to say that all trials were successes. Reichenbach's other cases would not interest Mill, for his rule deals only with the long run and provides no guidance for predicting the next instance.

I do not wish to enter the controversy over whether Reichenbach can justify the straight rule in preference to some 'crooked rule' containing an additional term which converges to zero as the number of trials approaches infinity, for I think the justification of any of these would share the same defect in Mill's eyes. Reichenbach is very anxious to preserve the force of Hume's sceptical conclusions, but he also believes that justification must contain some deductive element. This leads him, like Strawson, to appeal to an analytic statement. We know that in mathematics the very meaning of a statistical probability is the limit of a sequence of frequency ratios, and to say that the sequence converges simply *means* that for any given ϵ-neighbourhood of the limit p we can find some number N of trials such that beyond N the frequency ratios will remain within ϵ of p. It follows analytically, from the definitions, that if the sequence does converge we can get arbitrarily close to the real limit by using the straight rule, as long as we are willing to repeat the experiment sufficiently many times. If the sequence is finite, we will actually reach that limit.

---

[1] Reichenbach [1938], pp. 339–63.

[2] *Ibid.*, p. 350.

[3] Lenz [1960], p. 300.

For Reichenbach the great merit of this approach is that use of the straight rule is justified analytically. At no point do we make any assumptions about the overall character of the world which would have to be discovered empirically and inductively, thus leading to circularity. Nowhere do we make any *a priori* assumptions or use any data except the observed frequency ratios. While the convergence may not show up for a long time, so that we must be patient in applying the rule, it will eventually reward us with the desired result. However, this very fact leads to a difficulty which Reichenbach certainly recognises and which Mill would consider a crucial weakness. For it is always possible that the sequence we are investigating may not converge to any limit at all, no matter how long we keep experimenting. We have no *a priori* guarantee that such a limit exists, and thus the most that Reichenbach can and does claim is that if success is possible, *i.e.* if a limit really does exist, repeated application of his straight rule will eventually get one arbitrarily close to it.

Thus Reichenbach's justification is a purely hypothetical statement in which we never can be sure if the antecedent is true or false. And despite Reichenbach's calling this a pragmatic justification, it is certainly inadequate for the scientist. Without further information we simply have an interesting mathematical fact which is true in all possible worlds and thus gives us no guidance about how to act in this one. It is certain, but just for that reason useless. In making decisions, a scientist needs a categorical statement reflecting the peculiarities of this particular world, even if that statement is not certain. In so far as he decides to accept or use a particular inductive prediction, he must take the antecedent to be true, *i.e. assume* that the limit does exist. But if we now take uniformity in a somewhat broader sense which includes statistical uniformities (converging sequences) but rules out non-converging sequencies, we are right back at Mill's second claim for the significance of uniformity: it is that principle which is accepted in practice by anyone who uses an inductive argument. And the uniformity we have observed in nature means that in most cases (though by no means all) the limits do exist, and indeed the convergence is sufficiently rapid so that we can make confident predictions after relatively few trials. In a less uniform world

Reichenbach's method would be less valuable. One might, with tongue in cheek, call uniformity the *minor* premise that Reichenbach needs.

Finally, let us turn to Carnap's grandiose attempt to develop a fully quantitative inductive logic, and see how his ideas and Mill's might be used to clarify and supplement each other. Like Strawson, Carnap conceives of induction in much broader terms than do Mill or Reichenbach. His general problem may be phrased as follows: given any hypothesis H, and a body of evidence E, find the degree of confirmation $C(H, E) = p$ which E lends to H, where p is a real number such that $0 < p < 1$. This in fact includes all of deductive logic, with entailment and contradiction the special cases for which $p = 1$ or 0. But it is most conveniently applicable to a case like Mill's.

One must note that Carnap's method applies in full detail only to rather artificial and highly structured cases, like drawing coloured balls from an urn. One must assume, for example, that all the observed instances are fully comparable and interchangeable; that there is a limited and quite definite number of possible properties in the area of interest (*e.g.* colours) for the objects under study; and that each object clearly has one of these properties rather than another. In practice this can be achieved only by approximation or by somewhat arbitrary extra assumptions. Nevertheless, if we can get precise numerical results in these cases we may be able to give a qualitative account of the more general cases.

In the case of balls in an urn, we assume that each ball may have any one of exactly k colours. Our evidence E is that $(n-1)$ balls have been drawn so far, and that all of them have been found to have the same colour. (If we could limit the possible colours of crows in this way, we might apply the example to the $n-1$ crows we have so far observed, all of which were black.) Our hypothesis H is that the next, or nth ball drawn will also have that colour. Note that the evidence E is just what Mill was considering, but the hypothesis H is not the universal law that Mill wanted, *viz.* 'All balls will have that colour', but simply that the next will.

Lakatos has pointed out that this change reflects a serious weakness in Carnap's system.[1] By any of the 'natural' rules that Carnap considers, the degree of confirmation turns out to be zero if the reference class for a universal hypothesis is infinite. This might be less of a problem if we were dealing only with finite classes. Nevertheless, we would still have to know the exact size of the class to get numerical precision. This number will certainly be unknown for most scientifically interesting cases, and if we may observe the same object more than once (as with replacing balls in

[1] Lakatos [1968], Part 2, pp. 321–74.

the urn, rather than removing them permanently) the class is potentially infinite.

In order to overcome this difficulty, Carnap developed his 'qualified instance method', which concentrated on calculating the degree of confirmation for particular, rather than universal hypotheses. This represented a significant weakening of Carnap's original programme, since it greatly restricts the class of hypotheses to which his method applies. It seems justified only if theories involving universal laws are dispensable both as end products of scientific investigation and as ingredients in the logic of new discoveries, a thesis which is highly implausible and likely to stultify the growth of science.[1] But at least Carnap's approach of dealing only with the next case is in the spirit of Mill's inferring from particulars to other particulars, especially if the next case is truly random.

In an early paper, Carnap suggested that an inductive logic must be based on principles of symmetry or indifference. If we simply consider possible colour distributions among the totality of n balls, including the yet unexamined one, we have no *a priori* reason for believing that any one description of this collection is more likely than any other, so we assign them equal prior probabilities.[2] This general principle seems sound enough, and is closely related to the classical principle of sufficient reason. But the problem is how to draw up the various descriptions to which equal probabilities should be assigned. What Carnap here calls Method *I* uses individual descriptions, or microstates, *e.g.* 'first is red, second is white, third is blue, *etc.*'. Method *II* uses statistical distributions, or macrostates, *e.g.* 'four are red, two are white, three are blue, *etc.*' regardless of the order.

If we assign prior probabilities by these two methods, we can calculate the degrees of confirmation by relatively straightforward rules of the probability calculus. It turns out that by Method *I*, $p = C(H, E) = 1/k$, quite independent of the value of n. On the other hand, Method *II* gives $p = n/(n+k-1)$. Carnap then concludes that Method *II* is to be preferred, since it incorporates the principle of learning from experience.[3] Indeed, Method *II* leads directly to Laplace's rule of succession. Learning from experience would certainly seem to mean that the probability estimates should change with each new observation, but according to Method *I* they remain the same. No matter how many red balls have been observed, with no non-red balls, there is no more reason to expect that the next will be red than there was for the first.

However, this does not settle the case. There are many situations, like tossing an apparently unbiased coin and getting a long run of heads, where we would not want to change the probabilities. In particular Method *I* would seem appropriate whenever we believe we are dealing with a situation in which all events happen in complete independence of each other. In such a world there is no causality and no uniformity. Everything happens in a purely random fashion and order, and in the long run all possibilities in the area of investigation are realised approximately the same number of times. In dealing with each case we therefore have no reason for trying to apply our experience of previous cases, even if it seems to fall into patterns. In some restricted or artificial situations this attitude may be most appropriate. Nevertheless, a world of this sort would be a very nice model of the state of complete chaos in which induction is certainly not rational. And our past experience has shown that the real world has generally not been like that. Whether or not he could have survived in such a world, anyone who had bet on Nature according to Method *I* would have lost all his stake long ago. Method *I* is clearly unsatisfactory as a tool for most scientific investigation, even though we could only have discovered this fact empirically.

---

[1] Lakatos [1968], especially, pp. 330–42.

[2] Carnap [1955], and Carnap [1945].

[3] Carnap [1955], p. 277.

Even if we therefore accept Carnap's principle of learning from experience and incorporate normal induction in some way, there are still problems. For Methods *I* and *II* are not the only possible ways of assigning prior probabilities, and while Laplace's rule may be plausible, it is certainly not the only plausible rule. There would seem to be three reasonable intuitions about any choice of an inductive rule for the case at hand, where we want to represent p as a function of n. (*1*) For $n = 1$, so that we have had no prior experience but know the possible outcomes, $p = 1/k$. This incorporates our indifference principle. (*2*) p should be a monotonically increasing function of n. This incorporates the notion that each additional favourable instance should strengthen our belief or expectation. (*3*) p should approach *1* asymptotically as n approaches infinity, but should never reach it. This incorporates Hume's claim that we can never reach theoretical certainty by inductive procedures, but can at least reach a practical or 'moral' certainty, corresponding to sufficiently high probability. Laplace's rule meets these three conditions, but so do infinitely many others. Method *I* satisfies (*1*), but not (*2*) and (*3*).

Carnap eventually came to recognise this fact himself. In his final replies to objections as well as in several previous papers, he recognised that at most an inductive logic could restrict us to a whole continuum of inductive methods, where the members of that continuum were characterised by a single real number parameter λ.[1] λ could take on all positive values from 0 to $+\infty$. The different values would characterise a family of functions, but the actual determination of the proper value of λ was outside of the system of inductive logic. For the case at hand, Carnap calculated the degree of confirmation to be $p = [(n-1)+\lambda/k]/[(n-1)+\lambda]$.[1] Let us see what happens if we now substitute particular values of λ into this expression. (*1*) For $\lambda = \infty$, we get $p = 1/k$, equivalent to Method *I*. (*2*) For $\lambda = k$, we get $p = n/(n+k-1)$, equivalent to Method *II* and Laplace. (*3*) For $\lambda = 1$, we get $p = (1/n)(n-1+1/k)$. This is a rule that was considered in detail by Salmon and finally rejected as an alternative to Reichenbach's straight rule for predicting limit frequencies.[2] Since it was proposed in a different context and for a different reason, it may be unfair to consider it here as a possible degree of confirmation, but it does satisfy all the conditions above. (*4*) For $\lambda = 0$, p is undefined for $n = 1$, but $p = 1$ thereafter. This corresponds to the original Reichenbach rule, though subject to the same qualifications as Salmon's. We note that it violates all three of the above constraints, but along with $\lambda = \infty$ it represents the extreme or limit cases. We see that in general the rate of approach of p to 1 varies inversely as the value of λ.

---

[1] Carnap [1963], Part V; and Carnap [1952].

We have seen what role λ plays in classifying possible inductive rules. But what is the significance of λ itself and how are we to determine, estimate, or choose its value? I wish to claim that it plays three important and distinguishable but interconnected roles, and through this to show more precisely the relation between Mill's and Carnap's ideas.

A. Ontologically, λ provides what we have been looking for to answer the charge of vagueness. For it is a precise, quantitative, numerical, scalar measure of the degree of uniformity of nature. The value we choose represents our estimate of how uniform nature is, at least in the area under investigation. We may make a poor choice, but the ontologically correct value determines the best rule to use. This value can be discovered only inductively, for it reflects our actual world and would be different in other possible worlds. Increasing λ indicates decreasing uniformity, up to the limit of chaos. Decreasing λ represents increasing uniformity, up to the maximally uniform world where we could correctly make every inductive inference on the basis of only one observed instance.

B. Epistemologically, λ represents the relative importance that we attach to *a priori* and *a posteriori* factors in gaining knowledge. By hypothesis, we know k *a priori*, but n reflects the amount of our actual experience. The Method *I* extreme represents a world in which all knowledge is *a priori*, since we can not use our past experience in any way. The Reichenbachian extreme is a world in which all knowledge is *a posteriori*, since we do not allow our awareness of alternative possible outcomes to enter our probability estimates. All of the intermediate values, like Method *II* and Salmon's, represent a belief that some admixture of the two factors is required, whatever the appropriate ratio.

---

[1] Carnap [1963], p. 976.

[2] Salmon [1963].

C. Pragmatically, λ represents the relative importance that one attaches to accuracy and efficiency in making inductive predictions. In William James's language, it is the balance between the ideals of 'gaining truths' and 'avoiding falsehoods'.[1] A Jamesian impatient to make as many true predictions as quickly, easily, and efficiently as possible, and not greatly bothered by the chance of false predictions, will choose a low value of λ, with the Reichenbachian case representing extreme credulity. A Cliffordian wanting overwhelming evidence before hazarding a prediction and very fearful of error will choose a high value of λ, with the Method *I* case representing extreme scepticism. The scientific community might recommend a general standard beyond the whims of each individual, but this would still involve an element of human choice.

Putting this all together, we see that any justification of induction, or of a general inductive logic leading to specific rules of inference, must be based on non-logical facts about the constitution of the actual world and man's need to act in it. The world does have a degree of uniformity and if we knew just what it was in some *a priori* way we could use this to satisfy simultaneously the ideals of accuracy and efficiency in the best way possible. At the same time we could draw epistemological conclusions about the relative importance of *a priori* and *a posteriori* knowledge in science. Since we do not know the true value, we still can and implicitly do make estimates. These are based largely on our previous experience of particular uniformities. At the same time, they can be influenced by epistemological doctrines held on other grounds and by practical considerations of urgency and risk.

Of course, even this picture is oversimplified. We may find that nature is non-uniform in its uniformities, so that in different areas of investigation we may get better results by choosing different values of λ. In addition, the pragmatic component involves not only the general personality of the investigator, but also the specific context, and the values here may be constrained by moral and political principles. But in some form there is always an expectation, or at least a hope of some particular degree of uniformity.

In setting up a system of inductive logic originally, theoretical considerations of simplicity and intuitive intelligibility may be important for an approximation. These might lead one to choose Method *II* ($\lambda = k$)

---

[1] James [1896], especially pp. 99–102.

rather than some other alternative. But if inductive logic is to be an effective tool for scientific reasoning, these considerations must ultimately be less decisive than facts about the world. Since the probabilities involved are relations between a hypothesis and evidence that may be very skewed, we cannot test them directly or separately as predictions of long-run frequency ratios. But by reviewing the overall state of scientific progress, we can decide to reject as a whole the system that leads to them. If we find that we are simply getting the expected results over and over again, we may decide that we have been too cautious and that nature is even more strongly biased in favour of the more uniform macrostates, even if we can find no way to embody this in elegant indifference principles. On the other hand, if too many of our predictions have proved false, we may raise λ and conclude that the world is less uniform than we originally hoped. But our goal is always to become better attuned to the world that we live in.

Massachusetts Institute of Technology

REFERENCES

Carnap, R. [1945]: 'On Inductive Logic', *Philosophy of Science*, **12**, pp. 72–97.

Carnap, R. [1952]: *The Continuum of Inductive Methods.*

Carnap, R. [1955]: 'Statistical and Inductive Probability', in E. H. Madden (*ed.*) : *The Structure of Scientific Thought*, pp. 269–79.

Carnap, R. [1963]: 'Replies and Systematic Expositions', in P. A. Schilpp (*ed.*): *The Philosophy of Rudolf Carnap*, pp. 966–99.

Cohen, M. R. and Nagel, E. [1934]: *An Introduction to Logic and Scientific Method.*

Goodman, N. R. [1955]: *Fact, Fiction, and Forecast.*

James, W. [1896]: 'The Will to Believe', in *Essays in Pragmatism*, pp. 88–109.

Lakatos, I. [1968]: 'Changes in the Problem of Inductive Logic', in I. Lakatos (*ed.*): *The Problem of Inductive Logic*, pp. 315–417.

Lenz, J. W. [1960]: 'The Pragmatic Justification of Induction', in E. H. Madden (*ed.*): *The Structure of Scientific Thought*, pp. 299–303.

Madden, E. H. [1960]: 'The Riddle of Induction', in E. H. Madden (*ed.*): *The Structure of Scientific Thought*, pp. 287–93.

Mill, J. S. [1898]: *A System of Logic.* Eighth edition.

Nagel, E. [1950]: *John Stuart Mill's Philosophy of Scientific Method.*

Quine, W. V. O. [1963]: 'On Simple Theories of a Complex World', *Synthèse*, **15**, pp. 103–6.

Reichenbach, H. [1938]: *Experience and Prediction.*

Reichenbach, H. [1944]: *Philosophical Foundations of Quantum Mechanics.*

Salmon, W. C. [1963]: 'Inductive Inference', in B. Baumrin (*ed.*): *Philosophy of Science—The Delaware Seminar*, **2**.

Strawson, P. F. [1952]: *Introduction to Logical Theory.*

Brit. J. Phil. Sci. 25 (1974), 319–342 *Printed in Great Britain*

Discussions

POPPER AND QUANTUM LOGIC

This paper is *not* written in order to vindicate quantum logic. Rather it is written to rebut a mistaken criticism of quantum logic. The criticism is due to Popper ([1968]) and it concerns a classical paper by Birkhoff and von Neumann ([1936]). As is well known, this latter paper inaugurated the study of two rather different though interconnected kinds of problem, namely (*1*) the philosophical problem whether quantum mechanics involves a non-classical logic, the so-called *quantum logic*, and (*2*) the mathematical problem of characterising the basic structure of quantum mechanics, *i.e.* Hilbert space, by physically plausible axioms of a *lattice theoretical kind*. As was to be expected, the philosophical problem concerning the logic of quantum mechanics has remained controversial, whereas there has been considerable progress in solving the mathematical problem of axiomatisation (*cf.* Varadarajan [1968]). But until recently one thing has *never* been called in question, in either the philosopher's or the mathematician's camp, namely that Birkhoff and von Neumann's paper really points to well-defined problems. Only Popper had doubts in this respect. Having repeatedly read his paper of 1968, I can only summarise his criticism as saying, *horribile dictu*: in their paper of 1936 Birkhoff and von Neumann talked about something which really does not exist and the non-existence of which can, moreover, be shown by means of rather trivial mathematical considerations. Since this account may sound like a caricature of Popper's criticism, let me quote from his own summary, according to which 'Birkhoff and von Neumann's famous paper . . . culminates in a proposal which clashes with each of a number of assumptions made by the authors' (p. 682). Popper also says that Birkhoff and von Neumann's proposal 'breaks down because of some very trivial results of probability theory, or measure theory' (p. 683). Popper does not say that all later work founded on Birkhoff and von Neumann's paper has been done in vain. But he seems to think that the paper itself suffers from a fundamental inconsistency.

Until now, Popper's criticism seems to have received no official rebuttal. An explanation of this fact, if it is a fact, could very well be that no competent reader of Popper's paper has thought it worthwhile to answer it. If I now believe it advisable not to ignore his paper, it is because the paper is written by someone who is, perhaps, the most influential philosopher of our time. My article is divided into three parts corresponding to the three parts of Popper's paper. First, I give an outline of the main ideas developed in Birkhoff and von Neumann's paper (1). This makes my argument self-contained and replaces Popper's own outline which I find insufficient and misleading. Then I shall show that two mathematical arguments of Popper, valid as they are, rest on interpretative premises completely without foundation in the position actually held by Birkhoff and von Neumann (2). It thus will become evident that Popper misunderstood the authors' paper, and this claim will be confirmed by the discussion of some additional remarks made by Popper (3).

1 Birkhoff and von Neumann start their investigation (sections 2–6 of their paper) by contrasting two structures which can be associated with every physical

system. On the one side there is the *set of possible experimental propositions* that can be made about the system. Any element of this set is an n-tuple of measurements $M_1 \dots M_n$ together with a subset S of the observation-space of the M_r, *i.e.* of the space of 'ideal' outcomes $x_1 \dots x_n$ of the $M_1 \dots M_n$ respectively. The experimental proposition belonging to $M_1 \dots M_n$ and S is considered as the prediction 'that the point $(x_1 \dots x_n)$ determined by actually measuring $M_1 \dots M_n$ will lie in the subset S' (section 2). On the other side there is the *mathematical phase space* of the physical system. Any element of the phase space is a mathematical representative of the system, thereby representing the maximal knowledge that can be gained about it. Having at hand the two concepts of experimental proposition and phase space, Birkhoff and von Neumann establish a *fundamental connection* between them. This is done for classical mechanics and quantum mechanics separately, the general idea, however, being to associate with every experimental proposition as its mathematical representative the set of those points in phase space that correspond to states in which the given experimental proposition is true. In this way, as regards classical mechanics the mathematical representatives of experimental propositions are the Lebesgue-measurable subsets of classical phase space taken *modulo* sets with Lebesgue-measure 0, whereas in quantum mechanics the corresponding representatives are the closed linear subspaces of Hilbert space. This correlation between experimental propositions and their mathematical representatives in phase space is strengthened by the fact that certain natural relations and operations belonging to the mathematical representatives may mirror certain logical relations and operations concerning the experimental propositions. Thus in the classical case the representatives are ordered by set inclusion and the totality of them is closed under the operations of set-theoretical intersection, union and complement. Furthermore, in this case these relations and operations certainly correspond in a straightforward way to the logical relations and operations of implication, conjunction, disjunction and negation concerning experimental propositions. Likewise in quantum mechanics the closed subspaces of Hilbert space are ordered by set inclusion and their totality is closed under set-theoretical intersection, (topologically) closed linear sum and orthogonal complement. As distinguished from the classical case in quantum mechanics, however, it is not clear whether these relations and operations *generally* correspond to certain logical relations and operations concerning experimental propositions. Rather this question constitutes the main interpretative problem of so called quantum logic.

Birkhoff and von Neumann do not care much about this problem. Instead, in the second part of their paper (sections 7–14) they go on to a *comparison* of the two types of the structures set up in the first part: the classical and the quantum mechanical sets of mathematical representatives of experimental propositions provided with the aforementioned relations and operations. As regards the *common* properties, it is pointed out that both structures are orthocomplemented lattices. If the quantum mechanical case is restricted to finite-dimensional Hilbert spaces, then another common property is modularity.[1] On the other hand, the essential *difference* is brought out in considering the *distributive law*:

---

[1] It has to be noted that von Neumann toyed with the idea of founding quantum mechanics on continuous geometries, *cf.* section *15* of the paper under consideration as well as von Neumann [1937]. Thus modularity is considered as being common to classical and quantum mechanics also in a wider sense.

this holds in the classical case, yielding a Boolean algebra, whereas it breaks down in quantum mechanics. Having reached these results by means of axiomatic analysis the authors turn the tables by asking the following question: The classical structures being sufficiently characterised, as regards their logical aspect, as Boolean algebras, how can we correspondingly characterise the quantum mechanical structures among the orthocomplemented modular lattices? The answer to this question, given in the rest of the paper, is its main mathematical achievement. It belongs to a series of investigations pursued by von Neumann and others since 1927 and sometimes called the von Neumann programme.[1] The solution given in his [1936] is as follows. Confining the investigation to the finite-dimensional case Birkhoff and von Neumann impose one further condition on the orthocomplemented modular lattices, namely irreducibility, which thus replaces classical distributivity. The physical meaning of irreducibility is that only the trivial experimental propositions are compatible with all experimental propositions. They then prove that these lattices are the lattices naturally attached to a finite-dimensional vector-space over a (not necessarily commutative) field, provided with a definite Hermitian form. Thus this solution of the characterisation problem is successful in as much as it yields the relevant algebraic structures. But it does not succeed in singling out the field of complex numbers as it is used in ordinary quantum mechanics.

2 Having outlined what is usually taken to be the main content of Birkhoff and von Neumann's paper I now turn to Popper's criticism of it. According to his own outline of the paper, Birkhoff and von Neumann proposed 'that we should base the (non-classical) probability calculus to be used in quantum mechanics on a non-Boolean algebra, by discarding the law of distribution' or rather '[by replacing it] by a weaker law, the so-called "modular identity" ' (p. 682*f*). Already this summary is misleading in at least two respects. First, as I shall argue in more detail below, probability considerations are not of any importance in Birkhoff and von Neumann's paper. Secondly, the distributive laws are *not* replaced by modularity, modularity belonging to that part of the axiomatics which is *common* to classical and quantum mechanics (*cf. above*, p. 320, footnote 1). Rather they are replaced by the axiom of *irreducibility* which alone brings into the axiom system a feature specific to quantum mechanics. I shall come back to this point, too (*cf. below*, 3). At any rate postulating irreducibility means denying the distributive laws, and Popper's main assertion in the second part of his paper is that this proposal of Birkhoff and von Neumann 'clashes with each of a number of assumptions made by the authors' (p. 682).

His *first argument* intended to demonstrate this clash runs as follows. The authors decidedly want to base quantum mechanics on an orthocomplemented modular lattice L. Furthermore, they tacitly or unconsciously assume (or are forced to assume by what they say in other places or, finally, by inaccurate way of expression encourage the reader to take them as assuming) that, in addition, L is *uniquely complemented*. But, curiously enough, the very same authors at about the same time when they published their paper on quantum mechanics, or at any rate only a little later, proved (and published their proofs) that all their assumptions on L taken together force L to be a *Boolean* algebra. Therefore their proposal to make L non-Boolean breaks down according to their own results. This is Popper's first argument, and as an argument it *is* quite valid. Unfortunately not all of its premises are true. This is certainly the case with regard to the assumption that L be uniquely complemented: Birkhoff and von Neumann did not dream of assuming that their lattice L is uniquely complemented simply because they *certainly knew* the trivial fact that the lattice of subspaces of Hilbert space does not have this property and, as has been recapitulated in the outline of their paper given above, the postulates about L were *meant* to characterise just *this* lattice. Likewise, the authors *certainly knew* that unique complementedness is incompatible with their axiom of irreducibility (which is constantly ignored by Popper). This is evident from von Neumann's lectures on continuous geometry given in the academic year 1935–6: There irreducibility is characterised as stating that apart from the two neutral elements of a lattice *no* element has a unique complement (*cf.* v. Neumann [1960], p. 3). Even Popper refers to these lectures as containing one of the above mentioned results, making the proposal of the 1936 paper impossible. How then, one is inclined to ask, can Popper believe that the critical assumption really is made in this paper?

---

[1] Papers up to 1936: Hilbert, von Neumann and Nordheim [1927]; von Neumann [1927] and [1932], Ch. IV; Jordan, von Neumann and Wigner [1934]; von Neumann [1936].

The only evidence he produces in favour of his imputation is that 'in their section 9, Birkhoff and von Neumann constantly speak of "the" complement and "the" operation of complementation' (p. 683). Now, in actually doing this Birkhoff and von Neumann in section *9* of their paper constantly speak of the 'complement' with respect to an *orthocomplemented* lattice in the sense of a structure in which, in addition to the two binary lattice operations, $\cap$ and $\cup$, a third unary operation '$'$' is introduced, meaning by the 'complement' of an element a the element a' uniquely assigned to *a by means of this operation.* They are cautious enough to warn the reader not to confound this procedure with the assumption that the lattice shall be uniquely complemented by repeatedly putting the word 'complement' in quotation marks. Also they explicitly mention as their two principal cases of application the (Boolean) lattices of subsets of a set with set-complementation as the operation '$'$' *and* the lattices of subspaces of a Hilbert space with the usual orthogonal 'complement' as the operation '$'$', this case being obviously an example of orthocomplemented lattices in which only the neutral elements have unique complements. Thus there is not the slightest reason to accuse them of assuming unique complementedness for their lattice, and Popper himself has to admit that 'they do not assert explicitly that their orthocomplemented lattice is uniquely complemented' (p. 683). Then why does Popper take them to make the assumption in question?

Apart from a (not very serious) misuse of the prime-notation a' in Birkhoff's book on lattice theory[1] (which is rightly criticised by Popper in a footnote (p. 683), but which does not occur in the paper in question) there is in the paper itself a slightly more subtle point which may have caused Popper's trouble about the uniqueness of complements. Not as part of his criticism, but already in his outline of Birkhoff and von Neumann's paper, Popper reports that the authors 'do not propose to alter the classical rules of negation or complementation. Thus they speak of an operation of complementation . . . , implying that negation is unique, and they say . . . that by the negation of an experimental proposition is

[1] Birkhoff [1948], p. 114*f.*

meant 'the experimental proposition corresponding to the set complement . . . ' which of course is a classical complement' (p. 682 *f*). Here we have possibly one other root of Popper's misunderstanding of Birkhoff and von Neumann's interpretation. He does not recognise that an orthocomplemented lattice, although not itself a Boolean lattice, may have ***Boolean sublattices*** and that in quantum mechanics every observation-space actually leads not only to one Boolean sublattice of the orthocomplemented lattice of all subspaces in Hilbert space, but also to a natural representation of this sublattice as a Boolean lattice of subsets of the given observation-space (*modulo* an ideal of negligible subsets). In these sublattices, orthocomplementation coincides with set-complementation (in observation-space) and is therefore unique not only in the sense of being a unique operation but also in the sense that ***within*** these Boolean sublattices there is really no other complement besides the one distinguished by orthocomplementation. Only outside the maximal Boolean sublattices are there complements different from the orthocomplement. Thus Birkhoff and von Neumann are perfectly justified in regarding the 'negative' of an experimental proposition as corresponding to the ***set-complement*** of the set corresponding to the given experimental proposition ***'in the same observation-space'*** as they cautiously add—a phrase suppressed in Popper's quotation. But, equally well, these circumstances may have caused Popper's imputation that Birkhoff and von Neumann assume unique complementedness even in the ***unrestricted*** sense, ***i.e.*** with regard to the whole lattice.

Be this as it may, Popper has a ***second argument*** against Birkhoff and von Neumann's proposal. Having finally admitted that the authors 'do not assert explicitly that their orthocomplemented lattice is uniquely complemented' he from now on tries 'to interpret or re-interpret [*sic*!] their proposal as follows: the lattice should be non-Boolean and orthocomplemented with a unique operation of orthocomplementation, but not with unique complements' (p. 683). In other words: Popper now tries to interpret Birkhoff and von Neumann as they certainly wanted to be interpreted and as everybody else—apart from Popper—certainly *has* interpreted them. 'But'—continues Popper—'the attempted re-interpretation breaks down . . . because of some very trivial results of probability theory . . . '. Let us now have a look at the relevance of these very trivial results of probability theory.

In this second argument Popper imputes to Birkhoff and von Neumann another assumption concerning their lattice L, in addition to L being modular and orthocomplemented. According to this assumption L is provided with a set M of real-valued functions satisfying a number of conditions *R1*–*R4*, the two most important being:

R2: If for $a, b \in L$ and for every $m \in M$ the equation $m(a) = m(b)$ holds, then $a = b$

R4: For every $m \in M$ and for every $a, b \in L$

$$m(a)+m(b) = m(a \cap b)+m(a \cup b).$$

Now Popper succeeds in proving that the new assumption taken together with the unquestionable assumptions on L once more leads back to the result that L is Boolean. Thus, a second breakdown of Birkhoff and von Neumann's proposal seems to be established.

It is not necessary to go into the details of Popper's proofs because they are in order. Like his first argument the second is quite valid: the new assumption imposed on L certainly forces L to be Boolean. But again this new assumption is Popper's and not Birkhoff and von Neumann's. It is *evident* from their paper that these probability assumptions are not part of the *explicit* assumptions on which the authors' attempted characterisation of Hilbert spaces is founded. On the contrary, their attempt could very well be considered as an attempt to answer the question: to what extent is it possible to characterise quantum mechanical phase space *without* bringing probability assumptions into play? The non-occurrence of such assumptions is as evident as is the non-occurrence of the assumption of unique complementedness in the first case. This is so, despite the fact that here, unlike in the first case, Popper does not admit that probability assumptions are not part of the explicit assumptions made by Birkhoff and von Neumann. Such an admission is absent simply because the second argument is meant to be the final blow to their proposal.

Thus, the only thing to be done is once more to ask why Popper could come to think that Birkhoff and von Neumann, if not explicitly then at any rate implicitly, made the probability assumptions imputed to them by Popper. In the twenty pages of their paper there are indeed two (and only two) short passages in which the word 'probability' occurs at all—once in each case. The *first* passage does not even belong to the postulational part II of the paper. Instead it belongs to part I (section *6*) in which the physical background is summarised. A statement about quantum mechanics is there made which, *if* it had belonged to the postulational part concerning the lattice L and the set M of real-valued functions, would have been worded as follows:

BN1: For $a, b \in L$ we have $a \leq b$ if and only if for every $m \in M$ the inequality $m(a) \leq m(b)$ holds.

The *second* passage does indeed belong to the postulational part (section *11*). Here it is stated that modularity could be inferred from, and in the finite-dimensional case even replaced by, 'the assumption that there *exists* a numerical dimension-function' on L (my italics). It is added that since the conditions defining a dimension function 'partially describe the formal properties of probability, the presence of modularity is closely related to the *existence* of an "*a priori thermodynamic* weight of states" ' (my italics). *Assimilating* once more what here is considered as a possible postulate on the above set-up of an additional set M of real-valued functions, we would get the postulate

BN2: There *exists* $m_0 \in M$ such that for every $a, b \in L$

(*i*) if $a \subset b$, then $m_0(a) < m_0(b)$

(*ii*) $m_0(a)+m_0(b) = m_0(a \cap b)+m_0(a \cup b)$.

Before coming back to Popper's paper it has to be emphasised once more that the postulates *BN1* and *BN2*, as they stand, are *not* part of Birkhoff and von Neumann's axiomatics intended to characterise the quantum mechanical phase space. Rather I have assimilated the two casual remarks actually made by Birkhoff and von Neumann to Popper's set-up of probabilistic assumptions allegedly made by them only in order to do justice to Popper as far as possible. Indeed, the two passages in question are the passages to which Popper refers as being the evidence that Birkhoff and von Neumann at least implicitly made the assumptions reconstructed by him as his *R1–R4* and especially the above quoted *R2* and *R4*. But the most one can make of the authors two probability remarks by transforming them into postulates regarding probabilities is seen by looking at *BN1* and *2*. Popper's *R2* (and also its converse *R1*) follows from *BN1*, and is even a weakening of it. But instead of leaving *BN2* in its ***existential*** form as it occurs in the casual remark of Birkhoff and von Neumann, Popper postulates that ***every*** function in *M* should have the property given by the equation in *R4* and in *BN2.ii*). In this way, *R4* combined with *R2* leads to a ***very strong*** axiom system of probability and not to an 'extremely weak' one as Popper says (p. 684). Popper himself should have been aware of this in view of his own proof that *R1–R4* force the lattice *L* to be a Boolean algebra. Obviously nothing of the sort follows from *BN1* and *BN2*, these two axioms being trivially satisfied by, for instance, the set of all quantum mechanical probability functions on a finite dimensional Hilbert space.

It now becomes evident that Popper ignores several points which certainly were known to Birkhoff and von Neumann already in 1936. For one thing, as was already displayed above, they allude to the dimension function as being an ***a priori thermodynamic*** weight of state, thereby reminding the reader that this highly degenerate function represents, both in classical and in quantum mechanics, the exceptional state of affairs, in which nothing is known of a given physical system as regards its contingent measurable quantities. ***This*** function happens to satisfy the equation in *R4*. Completely ignoring this hint of Birkhoff and von Neumann, Popper seems to think that once probability functions come up, ***every*** one of them is bound to satisfy this equation. Saying that his axiom system is extremely weak, *R4* seems to be for him so natural an assumption that he imputes it to Birkhoff and von Neumann without further ado. Indeed, Popper does not bring forward the axioms *R1–4* in order to argue: Birkhoff and von Neumann accept these axioms, but this makes their proposal impossible, and therefore they have to look for some other axioms. Rather, by distorting the assumptions they really though only incidentally made, he seems to think: Birkhoff and von Neumann simply ***must*** have assumed *R1–4* because these are the reasonable axioms about probabilities. But, and this leads to a second point, Birkhoff and von Neumann knew, what again Popper overlooks, that for projective and continuous geometries there exists one and only one (normalised) function satisfying the equation in *R4*: Popper could have taken this information from von Neumann's book ([1960], p. 58*f*), to which he himself refers in his paper. Accordingly, the similar and, in the finite dimensional case, even identical structure of quantum mechanics to that of projective and continuous geometries would have deterred them from accepting *R2* and *R4* in an axiom system defining the possible probability functions. As a consequence of *R4* the set *M* would consist of essentially one function which, in a Hilbert space, is clearly incompatible with *R2*. Thirdly, it is quite clear what Birkhoff and von Neumann would have postulated for the set *M* of possible probability functions instead of *R4* if they had intended a probabilistic characterisation at all. From a characterisation of the quantum mechanical probability functions actually given by von Neumann already in his [1927] and again in his famous [1932] book (ch. IV) it immediately follows that *R4* is valid if the equation therein is restricted to ***compatible a*** and *b*, ***i.e.*** to experimental propositions which can be tested simultaneously. Thus they would have postulated *R4* in this restricted form, but again Popper does not seem to be aware of this way out.[1]

3 At the beginning of the third part of his paper Popper interprets the proposal of Birkhoff and von Neumann as being a 'shear weakening' of the classical propositional calculus and he therefore finds it 'scarcely understandable when they assert . . . that "quantum mechanics has a *greater logical coherence* than classical mechanics"—unless by "greater logical coherence" they mean no more than greater logical weakness' (p. 684). Now, although the first part of this remark can be turned into an interesting problem, to which I shall presently return, the second part reveals once more that Popper simply did not understand Birkhoff and von Neumann's paper. By attributing to quantum mechanics a greater logical coherence as compared with classical mechanics, Birkhoff and von Neumann point to *the one* axiom on account of which the 'logic' of quantum mechanics is *not* just a weakening of classical logic: namely, to the axiom of *irreducibility*, which states that apart from the two neutral elements there are no elements in the lattice L which are compatible with all elements of L (section *12*). Moreover, the authors leave no doubt about the meaning of the phrase 'greater logical coherence' by saying that this greater logical coherence of quantum mechanics is indicated by their suggestion that 'quantum theory involves *irreducible* propositional calculi of unbounded complexity' (section *16*, my italics), irreducibility being defined in their section *12* as mentioned previously. Obviously, this axiom does not only negate the Boolean structure of L according to which every element of L is compatible with all elements of L. Rather it postulates the opposite case in which, apart from the neutral elements, *no* element of L has this property. Thus the complete axiomatic system of Birkhoff and von Neumann, far from merely weakening classical logic, conflicts with it in a particularly strong way.[2]

Actually, there is one interesting problem connected with the axiom of irreducibility in so far as only this axiom goes beyond the mere weakening of classical logic. Let L_g be the first order language (with equality symbol) whose non-logical symbols are the symbols for the usual lattice-theoretical operations $\cap$, $\cup$ and $'$ and the constants for the neutral elements. Let T_{qu} be the set of formulas of L_g valid in all orthocomplemented lattices derived from complex Hilbert spaces in the usual way and in particular being their *full* sub-space lattices. Let T_u be the set of formulas of L_g valid in all Boolean lattices. Finally, let T^*_{qu} (resp. T^*_u) be the set of all *quantifier-free* formulas of T_{qu} (resp. T_u). The problem is: is T^*_{qu} a subset of T^*_u? This is a non-trivial problem because, as is well known, T_u with L_g as its language is an open theory, *i.e.* a theory which can be axiomatised by a set of quantifier-free formulas. On the other hand, T_{qu} is not an open theory—the axiom of irreducibility belongs to T_{qu}. Therefore there are substructures of models of T_{qu} which are not sub-models. To the best of my knowledge, the problem just stated has not yet been solved. It has not even been seriously raised. But its solution would be interesting for the following reason:

---

[1] The characterisation alluded to a moment ago presupposes Hilbert space and pertains to the set of all quantum mechanical probability functions. A characterisation *ab ovo* and introducing only a (relative) *a priori* probability function is given by von Neumann in his [1937].

[2] There is, of course, the trivial exception of the Boolean algebra consisting of only two elements.

if it should turn out, as I would conjecture, that T^*_{qu} is a subset of T^*_{u}, then quantum logic conceived as a *propositional calculus in the usual sense* would really be only a weakening of the classical propositional calculus (as distinguished from 'quantum logic' which includes Birkhoff and von Neumann's axiom of irreducibility). In other words: the specific features of quantum logic would be displayed only by establishing this logic as an extended propositional calculus using quantifiers applied to propositional variables.

Concluding his paper, Popper asks, what reasons Birkhoff and von Neumann could have had to suspect that the logic of quantum mechanics deviates from classical logic, especially with regard to the breakdown of the distributive laws. He finds two reasons indicated by the authors. The first is that distributivity of L follows from the additional assumption that any two elements of L are compatible (section 10). It is true that Birkhoff and von Neumann give a really obscure re-statement of this assertion immediately following it. But Popper's main complaint is that they do not prove their assertion in the original version. Presumably he does not see that they have a technical definition of compatibility in their section 12 and that therefore their statement is meant to be a purely mathematical one, the proof of which they know but do not want to give in this paper. Actually it can be given in a few lines (*cf.* Varadarajan [1968], lemma 6.10), and there is therefore no reason for complaint.

Coming to the second reason why Birkhoff and von Neumann believe the distributive law to break down in quantum mechanics, Popper for the first time in his paper touches on a really serious problem concerning the interpretation of quantum mechanics. Also it has to be admitted that the passage to which Popper refers in this connection is very ambiguous. The authors here want to show that the distributive law is invalid by giving a physical illustration (section *10*). Because of the weakness of this passage in Birkhoff and von Neumann's paper I do not want to discuss Popper's objection to it. Rather I will conclude this article by reminding him of a different and wellknown illustration of the invalidity of the distributive law in quantum mechanics. Let us consider a one dimensional quantum mechanical particle having position coordinate x and momentum coordinate p. Let a (resp. a') be the experimental proposition corresponding to a measurement of position and the subset $x > 0$ (resp. $x < 0$) of the observation-space belonging to the position coordinate. Finally let b be the experimental proposition corresponding to a measurement of momentum and the subset $p > 0$ of the observation-space belonging to the momentum coordinate. Using the *generally accepted* representation of these three propositions by subspaces of the Hilbert space of the particle (*cf. above*, 1) it follows that

$$b \cap (a \cup a') = b \supset 0 = (b \cap a) \cup (b \cap a'),$$

showing the invalidity of the distributive law. There is one (and only one) non-trivial step in the calculation leading to this result: In showing that $b \cap a = b \cap a' = 0$ one has to use a theorem on Fourier transforms. Correspondingly there is one (and only one) problematic step in the physical interpretation accompanying this calculation: the formation of $b \cap a$ and $b \cap a'$, being the only ones involved in the calculation in which *incompatible* propositions are combined and, therefore, being the only ones (concerning the present case, of course) for which the interpretation of a quantum mechanical lattice operation is doubtful. Thus, in a certain sense the distributive law *does* break down in quantum mechanics, and there is, then, the important problem of an adequate interpretation of this result. Accordingly Birkhoff and von Neumann are right in their ***assertion*** about the distributive law. The ***interpretation*** of their result does not any more belong to the aims pursued in their paper: They explicitly conclude it with the question: 'What experimental meaning can one attach to the meet and join of two given experimental propositions?'

With regard to Birkhoff and von Neumann's 'thought experiment' Popper says at the end of his paper in a patronising manner that it 'is no more than a simple slip—one of those slips which, once in a lifetime, may happen even to the greatest mathematician' (p. 685). No doubt, Popper will be delighted to learn that he was wrong.[1]

ERHARD SCHEIBE
University of Göttingen

REFERENCES

BIRKHOFF, G. [1948]: *Lattice Theory.*

BIRKHOFF, G. and NEUMANN, J. VON [1936]: 'The Logic of Quantum Mechanics', *Annals of Mathematics*, **37**, pp. 823–43. Reprinted in Taub (*ed.*): *Von Neumann, Collected Works*, **4**, 1962, pp. 105–25.

HILBERT, D., NEUMANN, J. VON and NORDHEIM, L. [1927]: 'Über die Grundlagen der Quantenmechanik', *Mathematische Annalen*, **98**, pp. 1–30. Reprinted in Taub (*ed.*): *Von Neumann, Collected Works*, **1**, (1961), pp. 104–33.

JORDAN, P., NEUMANN, J. VON and WIGNER, E. P. [1934]: 'On an algebraic generalization of the quantum mechanical formalism', *Annals of Mathematics*, **35**, pp. 29–64. Reprinted in Taub (*ed.*): *Von Neumann, Collected Works*, **2**, 1961, pp. 409–44.

NEUMANN, J. VON [1927]: 'Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik', *Nachrichten von der Akademie der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, pp. 245–72. Reprinted in Taub (*ed.*): *Von Neumann, Collected Works*, **1**, 1961, pp. 208–35.

NEUMANN, J. VON [1932]: *Mathematische Grundlagen der Quantenmechanik.* English translation: *Mathematical Foundations of Quantum Mechanics*, 1955.

NEUMANN, J. VON [1936]: 'On an algebraic generalization of the quantum mechanical formalism (Part I)', *Mathematicheskii Sbornik*, **1**, pp. 415–84. Reprinted in Taub (*ed.*): *Von Neumann, Collected Works*, **3**, 1961, pp. 492–561.

NEUMANN, J. VON [1937]: 'Continuous Geometries with a Transition Probability', *unpublished manuscript*, reviewed by I. Halperin in Taub (*ed.*): *Von Neumann, Collected Works*, **4**, 1962, pp. 191–4.

NEUMANN, J. VON [1960]: *Continuous Geometry.*

POPPER, K. R. [1968]: 'Birkhoff and von Neumann's Interpretation of Quantum Mechanics', *Nature*, **219**, pp. 682–5.

VARADARAJAN, V. S. [1968]: *Geometry of Quantum Theory*, **1**.

---

[1] [*Added in proofs*]: When this note was already in print Professor Lakatos came across an exchange of letters between Popper and Lenk published in *Conceptus*, **3**, 1969, pp. 43-6, and concerning Popper's first argument (*cf. above*, pp. 321–323). This discussion, however, does not clarify the matter, for two reasons: (*1*) Lenk does not make it sufficiently clear of what he accuses Popper. (*2*) In trying to exemplify his objection in Hilbert space Lenk makes a rather elementary mistake.

QUANTUM FACT AND FICTION

In 1965 Professor Alfred Landé presented seven axioms from which certain quantum properties could be derived.[1] In the same year Abner Shimony constructed a model consistent with these axioms but with no quantum properties.[2] Thus it is certain that these original seven axioms do not imply quantum properties. In 1969 Professor Landé modified his 'previous incomplete derivation'.[3] The modification consists of leaving the first six axioms unchanged and completely rewriting the seventh axiom. In the following note the new seventh axiom is discussed.

The six axioms which Landé leaves unchanged characterise a class of probability matrices; the matrix element '$P(ri,sj)$' is the probability of a system passing the '(sj) filter' after having passed the '(ri) filter'. The filters orthogonal to the (ri) filter are those with the same first index (r) and different second indices (second indices other than i). This means that after a system has passed the (ri) filter it will certainly not pass a filter with the same first index and a different second index. The notation '$P(ri,sj)$' will be used below in some cases for the (i,j) element of the (r,s) matrix and in other cases for the (r,s) matrix itself. The context will make it clear whether reference is being made to a matrix element or to a matrix.

In his paper of 1969 Professor Landé uses the new axiom to derive the property of 'interference' for his postulated probabilities. The logical order of the steps in this derivation is as follows. First Landé assumes the new seventh axiom:

> (A) There is a relation between the matrices $P(ri,sj)$, $P(sj,tk)$, and $P(ri,tk)$. This relation is such that the last of these three is restricted when the first two are specified. This relation is also to 'coincide with the ordinary probability relation law of classical physics in the average'.[4]

The nature of the 'average' is not specified in the axiom.

The second step is the following:

> (B) Write the elements of the P matrices as the absolute squares of complex quantities ψ which have unspecified phases.

This can always be done since the elements of the P matrices are all real and positive. The third step is to make the following assumption:

> (C) The complex ψ obey matrix multiplication.

Landé argues that this assumption is justified because it has an implication which fulfills the condition (A) of the new axiom. What Landé claims to be this implication of (C) is that:

> (D) There is a relation between the matrices $P(ri,sj)$, $P(sj,tk)$, and $P(ri,tk)$ which becomes
>
> $$P(ri,tk) = \sum_j P(ri,sj)\, P(sj,tk) \qquad (1)$$

when an average is taken over the possible phases that the ψ might have.

But this implication which Landé draws from assumption (C) can not be a fulfilment of the conditions of the new axiom unless two additional assumptions are made:

---

[1] Landé [1965].

[2] Shimony [1966].

[3] Landé [1969].

[4] *Ibid.*

> (*E*) Expression (1) above is the 'ordinary probability relation law of classical physics'.
>
> (*F*) The average which leads to the 'law of classical physics' is the phase average introduced in (*D*) above.

A second implication of assumption (*C*) is that the postulated probabilities have the desired property of 'interference'.

In order to derive interference as an implication of the new axiom it would have to be shown that assumption (*C*) is necessary to the fulfilment of the conditions of the new axiom. This would mean that there exists not a single relation between the matrices $P(ri,sj)$, $P(sj,tk)$, and $P(ri,tk)$ which will become (1) under any conceivable average without implying (*C*). This would be a strong theorem. It is not proved rigorously in Landé's paper. Without this proof we can not say that the new axiom implies interference of probabilities. Is assumption (*C*) sufficient to fulfil the conditions of the new axiom? Not without the additional assumptions (*E*) and (*F*). In assumption (*E*) the Markoffian processes characterised by (1) are singled out as the 'ordinary' probabilistic processes of classical physics. This needs more justification than Landé has given. What is it that rules out the classical processes which do not obey expression (1)? Assumption (*F*) requires that a particular phase average should lead to certain classical properties. If this phase average were a general prescription for recovering classical properties from Landé's axioms then its role as a 'classical limit' would have to be explained. But it will be shown below that the phase average plays this role only in the new axiom. If it is applied to other of the axioms an impossible set of classical properties is implied. This makes the role of the phase average seem especially puzzling.

Actually, even if we make assumptions (*E*) and (*F*) there is a difficulty in deriving expression (1) from assumption (*C*). To see this we turn next to the mathematical aspects of Landé's derivation. In step (*B*) above the *P* matrices are written, element by element, as the absolute squares of certain complex quantities with unspecified phases. We can define these quantities by writing

$$P(ri,tk) = \psi\,(ri,tk)\;\psi^{*}\,(ri,tk),$$

where we set

$$\psi(ri,tk) = R(ri,tk)\;e^{i\phi\,(ri,tk)},$$

and require that the *R*'s be real and that

$$P(ri,tk) = R^{2}(ri,tk),$$

element by element. In step (*C*) we make the very special assumption that the quantities ψ obey matrix multiplication. Mathematically this means that we can write

$$\psi(ri,tk) = \sum_{j} \psi(ri,sj)\;\psi(sj,tk).$$

From this we can derive an identity which we write as

$$P(ri,tk) = \left|\sum_{j} R(ri,sj)e^{i\phi\,(ri,sj)}\;R(sj,tk)e^{i\phi\,(sj,tk)}\right|^{2}$$

Taking the absolute square and using the definition of the ψ we can rewrite this identity as the sum of two terms,

$$P(ri,tk) = \sum_j P(ri,sj)\, P(sj,tk) + I,$$

where

$$I = \sum_{j \neq j'} R(ri,sj)\, R(ri,sj')\, R(sj,tk)\, R(sj',tk)$$
$$\cdot\, e^{i(\phi(ri,sj) + \phi(sj,tk) - \phi(ri,sj') - \phi(sj',tk))}.$$

Recall now that the phase angles Φ are all unspecified. If we average over all possible values of these angles the term 'I' vanishes *except* when $r=t$ and $i=k$. In this exceptional case we can write

$$I = \sum_{j \neq j'} P(ri,sj)\, P(ri,sj')$$

using a property which Landé derives from his axioms, namely that

$$\psi\,(ri,sj) = \psi^{*}(sj,ri).$$

Physically, this exceptional case is that in which a system passes the same filter twice. Putting this term into the identity we obtain

$$\begin{aligned} P(ri,ri) &= \sum_j P(ri,sj)\, P(sj,ri) + \sum_{j \neq j'} P(ri,sj)\, P(ri,sj') \\ &= \sum_j \sum_{j'} P(sj,ri)\, P(ri,sj') \\ &= 1, \end{aligned}$$

again drawing upon postulated properties of the matrices.

Because of the exceptional case noted above, we see that the actual implication of assumption (C) is not exactly (D) of Landé's argument. The actual implication is:

> (D') There is a relation between the matrices $P(ri,tk)$, $P(sj,tk)$, and $P(ri,tk)$, namely,
>
> $$P(ri,tk) = \sum_j P(ri,sj)\, P(sj,tk) + I$$

and this becomes

$$\begin{aligned} P(ri,tk) = {} & \sum_j P(ri,sj)\, P(sj,tk) \\ & + \delta_{rt}\delta_{ik} \sum_{j \neq j'} P(ri,sj)\, P(ri,sj') \end{aligned} \tag{1'}$$

when an average is taken over the possible phases that the ψ might have.

One way of remedying this discrepancy would be to modify assumption (C) so that (D), and not (D'), would be implied. Then the necessity of the modified assumption would have to be shown. Another remedy would be to change assumption (E) and assume that expression (1′), and not (1), is the 'ordinary probability relation law of classical physics'. Then this modified assumption would have to be made plausible.

It was mentioned above that the phase average can not be a general prescription for recovering classical properties from Landé's axioms. If it were such a prescription, then it could be applied to any of the other axioms to obtain some classical property. Among the properties inherent in Landé's other axioms is:

$$P(ri,ri') = \delta_{ii'}. \tag{2}$$

If we take the phase average of both sides of this expression it remains unchanged. If the phase average were a 'classical limit' then both (1), or (1'), and (2) would be properties of classical probabilities. This would, in turn, as Landé points out, imply that some of these probabilities are either negative or imaginary. Classical probabilities must be all real and positive and it follows that the role of the phase average can not be that of a general prescription, as stated above.

Another point should be touched briefly; Landé has based some of his arguments about quantum theory on the work of Epstein and Ehrenfest to extend the application of Duane's quantisation rule. This work is very interesting. But these authors concluded that Duane's view could not be tenable because no consonant account of coherence effects (like those in Fresnel diffraction) could be found.[1] If Landé is to advocate Duane's view then this objection should be discussed and answered if possible.

We are indebted to Professor Landé for bringing an important perspective to the discussion of quantum theory. A fuller consideration of the points which make his view unique could help to make his insights more generally understood. Answers to the questions raised here could be a step in this direction.

D. K. NARTONIS
Principia College, Illinois

REFERENCES

EHRENFEST, P. and EPSTEIN, P. [1924]: 'The Quantum Theory of Fraunhofer Diffraction', *Proceedings of the National Academy of Science*, **10**, pp. 133–46.

EHRENFEST, P. and EPSTEIN, P. [1927]: 'Remarks on the Quantum Theory of Diffraction', *Proceedings of the National Academy of Science*, **13**, pp. 400–16.

LANDÉ, A. [1965]: *New Foundations of Quantum Mechanics*.

LANDÉ, A. [1969]: 'Quantum Fact and Fiction, III', *American Journal of Physics*, **37**, pp. 541–3.

SHIMONY, A. [1966]: 'Basic Axioms of Microphysics: *review of Landé* [1965]', *Physics Today*, **19**, pp. 85–94.

ARE THE 'MEMBERS' OF BIOLOGICAL SPECIES 'SIMILAR' TO EACH OTHER?

Of the battery of objections raised by L. J. Cohen[2] to Stephen Toulmin's evolutionary account of scientific change,[3] I wish to object to the premises of one. Cohen discusses two fundamental flaws in the analogy which Toulmin draws between the evolution of biological species and rational disciplines. I will pass over the first which concerns the difference between 'coupled' and 'uncoupled' evolution. The second concerns the similarity required for individual organisms to be members of the same species and the lack of such similarity among the concepts which contribute to the intellectual content of a single rational discipline. Cohen argues:

---

[1] Ehrenfest and Epstein [1924] and [1927].

[2] Cohen [1973].

[3] Toulmin [1972].

> But a rational discipline is not merely a biological species: it is not a species at all—at any rate as Toulmin conceives it. It does not consist of a set of individual members that are similar to one another in all relevant respects and may be said to instantiate it.
>
> Again, concepts of physics are not, as such, all closely similar to one another, like the members of a species or population. Rather, they are almost all importantly different from one another, like the concept of an electron and the concept of a proton. Each such concept may conceivably have a few variant forms that are in competition with one another. But the great bulk of the 'concepts, methods and aims' within a rational discipline at any one time manage to be quite different from one another and yet not be in competition, but in systematic association, with one another.
>
> The relation between a concept and the rational discipline to which it belongs is, roughly, that of part to whole, not that of member to species or member to population.[1]

I agree with Cohen's observations about the relation between scientific concepts, though I'm not so sure about the relation which they have to scientific disciplines. But exactly the same observations can be made about individual organisms and the units of evolution which they comprise, units commonly called biological species or populations. Cohen's mistake is that he continues to treat biological species as if these evolutionary units were classes of similar organisms. On this score he is joined by nearly all scientists and philosophers not working directly in evolutionary biology.[2] But species as evolutionary units need not be and frequently are not comprised of similar individuals. For example, the various breeds, races, subspecies, *etc.*, which make up polytypic species exhibit as much dissimilarity as exists between most groups which are counted as separate species. The opposite situation exists for sibling species. Worse still, the various castes, sexes, *etc.*, which make up polymorphic species exhibit just as much dissimilarity. For example, queens, drones and neuter workers in some species of social insects differ from each other as much as do the concepts of proton, electron and neutron in atomic theory. In some polymorphic species, for instance, the male lives as a minute parasite on the female.

In the case of polytypic species, the various subspecies, *etc.*, frequently compete with each other; the various morphs in polymorphic species necessarily exist in systematic association. But of greatest relevance to Cohen's argument is that even when the members of a species are fairly similar to one another, they are *not* included in the same species *because* they possess any prerequisite degree of similarity. In attempting to distinguish species, degree of similarity and/or difference *does* play a role, just as IQ tests help in discovering how bright children are, but I take it to be a commonplace among philosophers that such methods of discernment do not define the relevant concept.

---

[1] Cohen [1973], pp. 48–9.

[2] For a traditional analysis of biological species as classes, see Buck and Hull [1966]; for a discussion of why this analysis will not do for the purposes of evolutionary biology, see Hull [1973] and [1974]; this position has been urged by Ernst Mayr for some time, Mayr [1969]; see also Ghiselin [1969].

Cohen's difficulty is that he has not made the conceptual shift necessary to view the units of biological evolution as 'populations'. On such an interpretation, the relation between individual organisms and their species is more like part to whole than member to class. Such a conceptual shift may introduce chaos into the principles of biological classification, but it causes no special problems for the formation of evolutionary theory. In fact, it may be necessary for an adequate understanding of evolutionary processes. I must add, however, that the notion of a biological population, not to mention the more generalized notion of population introduced by Toulmin, is far from crystal clear. Even so, it is still necessary to point out that what Cohen takes to be a difference between biological and conceptual evolution is actually a similarity, in fact a common difficulty.

DAVID L. HULL
The University of Wisconsin
Milwaukee

REFERENCES

Buck, R. C. and Hull, D. L. [1966]: 'The logical structure of the Linnaean hierarchy', *Systematic Zoology*, **15**, pp. 97–111.
Cohen, L. J. [1973]: 'Is the progress of science evolutionary?', *The British Journal for the Philosophy of Science*, **24**, pp. 41–61.
Ghiselin, M. [1969]: *The Triumph of the Darwinian Method.*
Hull, D. L. [1973]: 'A populational approach to scientific change', *Science*, **182**, pp. 1121–4.
Hull, D. L. [1974]: *The Philosophy of Biological Science.*
Mayr, E. [1969]: *Principles of Systematic Zoology.*
Toulmin, S. [1972]: *Human Understanding*, **1**.

PROFESSOR HULL AND THE EVOLUTION OF SCIENCE

Professor Hull has made a gallant attempt[1] to defend Toulmin's evolutionary theory of scientific development against one of the two criticisms urged in my [1973]. Even if the defence were successful on this point, my other criticism of Toulmin's theory—on the coupling-decoupling issue—would still be fatal to a Darwinian theory of science. But a brief reconsideration of the problem will show that Hull's defence does not succeed.

According to Hull 'Cohen's difficulty is that he has not made the conceptual shift necessary to view the units of biological evolution as populations'. In fact, however, I have discussed this conceptual shift, and related issues, in some detail elsewhere ([1962], section 14), and I did not resume the topic in my criticism of Toulmin's theory because I thought it would complicate the argument unnecessarily. No doubt a biological species can usefully be conceived not only as a class that has many individual organisms for members, but also as a populational whole that has these organisms for its parts and is an 'integrated unit of biological function' as Ghiselin ([1969], p. 90) calls it. The former conception has an obvious appropriateness for some taxonomic purposes, the latter for some explanatory ones. Even according to Ghiselin ([1969], p. 86)—one of Hull's authorities—both conceptions are equally legitimate. I chose to criticise Toulmin's theory in terms of the class conception for two reasons. First, whereas organisms can be regarded both as instances and as components of a species, 'the concepts, methods and aims' of a rational discipline can be regarded only as the components of such a discipline and not as its instances. Secondly, the class conception of species serves to highlight a crucial flaw in Toulmin's analogy between species and rational disciplines, while substantially the same disanalogy obtrudes itself (as I shall shortly show) even if species are conceived as integrated wholes.

---

[1] This *Journal*, pp. 332–334. I am grateful to Professor Hull for letting me see a copy of his typescript prior to publication.

Of course, whatever Darwin's own view were, so far as a species is conceived as a class in modern evolutionary biology, it is not so much a class of phenotypically similar organisms, but rather a class of organisms that are similar to one another in certain other respects that are causally more fundamental. The organisms may be taken to be similar to one another in respect of the gene-pool which they share. They may be taken to be similar to one another in respect of the type of organism with which they interbreed; or—more exactly—they may be taken to belong, in the standard case, to one or other of two sub-classes the members of each of which are similar to one another in breeding only with members of the other sub-class, and they may thus be taken to be all similar to one another in belonging either to the one such sex-class or to the other. And also the individual organisms may be taken to be similar to one another in their ecological niche. Many questions notoriously arise here. Which of these features should be regarded as defining conditions on the concept of a species, and which as associated facts? In what contexts, if any, is it important to draw that distinction? How precisely are unisexual species, polytypic species, *etc.* to be accommodated? And so on. But whatever be the correct solutions of these problems our ability to conceive a species as a class of organisms that are similar to one another in certain relevant respects is unaffected. Correspondingly biologists like Mayr, whom Hull cites as one of his authorities, do not hesitate, on appropriate occasions, to conceive of species as classes. Mayr writes ([1969], p. 28)

> What the taxonomist observes directly are individuals which he sorts into phena. On the basis of certain biological concepts and information, such as an awareness of the possibility of sexual dimorphism, growth, alternation of generations, nongenetic modifications of phenotype, *etc.*, he assigns the phena to populations, which in turn he classifies into taxa. The ranking of a taxon as subspecies, species, or genus by a taxonomist is based on inferences drawn from the available data.

So it is hardly a mistake, as Hull claims, for me to 'treat biological species as if these evolutionary units were classes of similar organisms'. It would certainly be a mistake to treat them as just being classes of *phenotypically* similar organisms. But that was not how I treated them in my [1973].

Now the point at issue is whether Toulmin's analogy between a species and a rational discipline can be sustained. I argued in my [1973] that the 'concepts, methods and aims' composing a rational discipline were not analogous to the organisms of a species or population, as Toulmin claimed they were in his [1972], p. 141, because while the organisms of a species or population had to be similar to one another in relevant respects the concepts, methods and aims of a rational discipline had all to be different from one another in their characteristic roles. This is the heart of the matter, and it is not affected by whether we conceive a species as a class or as the populational aggregate of the members of such a class. In the one case the individual organisms are instances of the species: in the other they compose it. But in both cases they must essentially be similar to one another in relevant respects. Hull points out that 'queens, drones and neuter workers in some species of social insects differ from each other as much as do the concepts of proton, electron and neutron in atomic theory'. But in such a species there are very many individual queens, very many individual drones, and very many neuter workers; whereas in any one version of atomic theory, if unequivocally expounded, there is only one concept of a proton, only one concept of an electron, and only one concept of a neutron. In order to rescue Toulmin's analogy here Hull would have to suppose the existence of a polymorphic species in which there was, essentially, only one specimen of each variety of each morph. Perhaps a biological species sometimes *becomes* like that. But it does so only in the last stages of extinction. A species as a whole, with its self-reproduction over time, is essentially composed of more than one specimen for each persisting variety of its morphs or types, and the unity of a biological species certainly requires the organisms that are the specimens of any one morph or type to be similar to one another in all relevant respects from one generation to the next.

Turn it how you will, the evolutionary model cannot provide the 'precise and strict' analysis of scientific development that Toulmin seeks ([1972], p. 134).

L. JONATHAN COHEN
Queen's College, Oxford

REFERENCES

COHEN, L. J. [1962]: *The Diversity of Meaning.*
COHEN, L. J. [1973]: 'Is the progress of science evolutionary?', *British Journal for the Philosophy of Science*, **24**, pp. 41–61.
GHISELIN, M. T. [1969]: *The Triumph of the Darwinian Method.*
MAYR, E. [1969]: *Principles of Systematic Zoology.*
TOULMIN, S. [1972]: *Human Understanding*, I.

DID EINSTEIN'S PROGRAMME SUPERSEDE LORENTZ'S?

Dr Zahar's contribution[1] on the origins and development of Relativity sheds a new light on many problems. In particular, he presents a balanced reappraisal of Lorentz's mighty contribution to electromagnetism and relativity theory. He suggests that Einstein's initial programme was based mainly on the need to reconcile Maxwell's equations with the Relativity Principle and shows how Lorentz's efforts towards this goal provided Einstein with the main motivation and ammunition towards the formulation of his 'Special Relativity'. It is of interest that John

---

[1] Zahar [1973].

Saunders[1] reached similar conclusions in his doctoral thesis on the genesis of Einstein's ideas. It is also of interest that since the constant '*c*' is associated with Maxwell's equations in a context independent of its light-velocity significance, the extension of the Relativity Principle to electromagnetic phenomena virtually requires that the value and constancy of *c* should apply equally with respect to all inertial reference frames. I believe that this theoretical requirement concerned Einstein's thoughts just as much as did the long-standing (by 1905) experimental evidence regarding measurements of the velocity of light. Hence the Light Principle (*P*2) is, after all, firmly based on complementary theoretical and experimental premises.

In the second part of his paper, Zahar presents an original standpoint on the reasons for the triumph of Einstein's programme over Lorentz's, and there is no doubt that he has made a valid and important contribution here. The spectacular mathematical and experimental developments advanced by his General Relativity put the seal on the triumph of Einstein's approach. However, as Zahar[2] himself, indicates, this approach had been applauded and further developed (*e.g.* by Planck, Minkowski, Sommerfeld, Langevin, *etc.*) by the scientific world long before 1915, and there are *many* reasons why this happened. Some of these fit into the explanations given by the Kuhnian, Whittaker and positivist approaches and some of them do not.

There are good reasons why Einstein's radical theory should have attracted the immediate interest of theoretical physicists. The following suggest themselves:

(a) Einstein offered an elegant, compact and self-consistent mathematical theory based on two general assumptions and a number of operational definitions (consistent with his assumptions) for his space and time co-ordinates, *etc.* Minkowski rendered the theory even more mathematically elegant, though thereby obscured its physical content and significance.

(*b*) Einstein disclosed the implications of his theory and of the Lorentz transformation far more comprehensively than Lorentz or even Poincaré (who was groping towards them) and thereby opened up a whole new radical view of mechanics, space and time, mass and energy, *etc.*, which was nevertheless not inconsistent with Newtonian mechanics seen as a limiting case.

(*c*) The lack of evidence for an ether suggested that the foundation of Lorentz's approach could not be substantiated. Einstein's approach appeared to make the ether unnecessary and was based on foundations having strong experimental and theoretical support.

(*d*) Lorentz's programme lacked the 'creative shift' which could lend his theory the universal significance which flowed from Einstein's broad assumptions and their bold development.

(*e*) Lorentz had a traditional attitude to the meaning of the co-ordinates of his transformation, and so was unable to grasp the significance of different viewpoints of an event associated with different inertial systems, the time-dilatation effect, relativity of simultaneity, *etc.*

Under the circumstances, it is not surprising that leading scientists, including Lorentz himself, soon recognised the advantages (not to be confused with 'superiority') of Einstein's approach and its fascinating implications.

However all this suggests a further question—does the triumph of Einstein's programme mean that Lorentz's approach was 'inferior', beyond redemption and without possibility of useful development? Was its role only to serve as a prelude to Einstein's revision of physics? Zahar suggests that this was by no means the case, that Lorentz's approach was not inferior but rather complementary to Einstein's, though it had reached a stage where further progress required a 'creative shift' which failed to emerge.

---

[1] Saunders [1972].

[2] *Ibid.*, pp. 237–41.

However, one might go even further than Zahar: Lorentz's programme had by no means outlived its usefulness but its further development depended on some new observations and concepts which were not yet 'in the air'; they were in fact hastened by the stimulation of Einstein's parallel approach. Firstly, at the beginning of the century there was no astronomical evidence for the existence of a universal basic reference frame for energy propagation; there is now[1]—partly as a result of Einstein's stimulation to cosmological theory and observation. Next, the notion that gravitational and *all* other forms of energy shared the limiting velocity of light had not yet been accepted and awaited Einstein's universal approach for its exposition. And finally, Einstein's approach suggested that an observable absolute effect could arise as a result of uniform motion, and that this effect (time-dilatation) was a function only of the uniform velocity relative to a given inertial reference frame. It took fifty years of controversy for the logic of this consequence to become clear. In 1958, Geoffrey Builder,[2] in reply to the searching scepticism of Herbert Dingle, asserted that the existence of this absolute effect (an experimentally confirmed and theoretically valid consequence of Einstein's assumptions) 'demands our recognition of the causal significance of absolute velocities'; and he went on to show[3] that time-dilatation, and all other phenomena and concepts of Special Relativity, have an unambiguous meaning if, and only if, there exists a basic reference frame for light-propagation, so that relative to this frame uniform motion has an absolute significance and is associated with absolute effects due to the anisotropy of energy propagation relative to 'moving' bodies. Thus Einstein's oft-debated out-and-return journey assertion (in his 1905 paper) provides a clue to the existence of an 'ether' sought-after by Lorentz. And this ether, a universal reference frame for light-propagation, can now be strictly defined and observed astronomically.[4] The recognition that all forms of energy have the same velocity relative to the same basic reference frame provides the 'creative shift' whereby *all* field manifestations must be affected similarly by the anisotropy effect (*i.e.* in relation to moving bodies), so that all such fields have a common characteristic derived from the broader concept of energy associated with a universal basic reference frame.

The above arguments have, of course, already appeared explicitly or implicitly in a number of publications.[5] However, Zahar's discussion on Lorentz's developing apparoch enables one to see more clearly the power and limitations of the Lorentz programme and to pinpoint the 'creative shift' which could only come when theory and observation had developed another fifty years.

The wider picture that now emerges requires the recognition of two valid and complementary ways of formulating scientific theories, leading to two types of such theories. Following Einstein[6] and Saunders[7], we may distinguish between

---

[1] *E.g.* see Prokhovnik [1973].

[2] Builder [1958*a*].

[3] Builder [1958*b*].

[4] See note 1, above.

[5] *E.g.* see Prokhovnik [1963] and [1967].

[6] Einstein [1919].

[7] *Ibid.*, pp. 25–40.

'constructive theories' which attempt to interpret complex phenomena out of relatively simple assumptions (not necessarily based on direct observation), and 'theories of principle' which attempt to deduce the consequences from general principles evident in nature. The Kinetic Theory of Gases is a good example of a constructive theory, and Thermodynamics is a typical theory of principle. Thus also, Lorentz's approach aimed at a constructive theory, whereas Einstein consciously formulated Special Relativity as a theory of principle, realising correctly that this was *at that time* the only way to make substantial progress towards reconciling electromagnetic theory as described by Maxwell's Equations with the Principle of Relativity.

However, the progress of observations and ideas since 1905 has made it possible to now further develop Lorentz's approach thereby adding new meaning to Special Relativity, resolving its apparent paradoxes (*e.g.* how can the measure of light-velocity be the same for all inertial reference frames?), and linking it firmly with modern cosmological and gravitational theory[1]. It is suggested that each of these approaches has its place in the development of our understanding of nature. Depending on the circumstances, at some times one approach is the more appropriate, at other times the second is the more fruitful.

Finally, a mild rejoinder. Zahar backs his interpretation of the early Einstein's philosophical attitudes with statements by Einstein from his later years. I am not convinced that hindsight statements by Einstein (or anyone else) really throw much light on his attitudes and reasoning in his early years, say up to 1905. You can quote Einstein, like the Bible or Shakespeare, to prove almost anything, and in fact this is done by the various contributors to 'Schilpp's book on Einstein. I believe, with John Saunders[2], that Einstein's attitudes changed considerably during his lifetime—which is not remarkable since this happens with most thinking people—and hence his later views on his earlier work were bound to be coloured by his more mature attitudes. Indeed, Einstein himself was well aware[3] of this problem. Thus, for example, his conviction that mathematics alone can provide a priori information about the world was not evident in 1905, but appeared to develop with Minkowski's geometrical representation of his theory and the consequent elaboration and generalisation of the metric to describe gravitational phenomena. His subsequent efforts to find a unified field theory in the same way were overtaken by events when fields and fundamental particles were found to exist in apparently great variety and number. To be fair to Zahar, his quotations of Einstein's later thoughts and opinions are employed cautiously and intelligently: however, in less able hands, this practice can easily lead to quite specious conclusions.

S. J. PROKHOVNIK
University of New South Wales

REFERENCES

BUILDER, G. [1958*a*]: 'Ether and Relativity', *Australian Journal of Physics*, **11**, pp. 279–97.

BUILDER, G. [1958*b*]: 'The Constancy of the Velocity of Light', *Australian Journal of Physics*, **11**, pp. 457–80.

---

[1] Prokhovnik [1974].

[2] *Ibid.*, pp. 23–4.

[3] 'Every reminiscence is coloured by today's being what it is, and therefore by a deceptive point of view . . .', in Einstein [1951].

Einstein, A. [1919]: 'My Theory', *The Times*, 28 November, p. 13: reprinted in Einstein [1950].

Einstein, A. [1950]: *Out of my Later Years.*

Einstein, A. [1951]: 'Autobiographical Notes', in P. A. Schilpp (*ed.*): *Albert Einstein, Philosopher Scientist*, p. 1.

Prokhovnik, S. J. [1963]: 'The Case for an Aether', *British Journal for the Philosophy of Science*, **14**, pp. 195–207.

Prokhovnik, S. J. [1967]: *The Logic of Special Relativity.*

Prokhovnik, S. J. [1973]: 'Cosmology versus Relativity—the Reference Frame Paradox', *Foundations of Physics*, **3**, pp. 351–8.

Prokhovnik, S. J. [1974]: 'Is a Cosmological Substratum Compatible with Relativity', *forthcoming* in *The International Journal of Theoretical Physics.*

Saunders, J. [1972]: *Relativity and Ether.* (Doctoral Dissertation, University of New South Wales.)

Zahar, E. G. [1973]: 'Why did Einstein's Programme supersede Lorentz's?', *British Journal for the Philosophy of Science*, **24**, pp. 95–123 and pp. 223–62.

A LITTLE LEARNING CAN BE DANGEROUS*

It has been proved that, under certain assumptions, it pays you 'in expectation' to acquire new information, when it is free. A precise formulation of this thesis, together with a proof, was given by Good,[1] with historical references to Carnap, Ayer, Raiffa and Schlaifer, and Lindley. The 'expectation' in this result is *your own* expectation. In the present note it will be pointed out that the result can break down when the expectation is computed by someone else. It is of course familiar that an experiment can be misleading by bad luck, but this is not by itself a justification for ignoring free information. It is perhaps surprising at first sight that, 'in expectation', an experiment can be of negative value to a subject, in the opinion of *another* person who may or may not be better informed or have better judgment.

The argument supporting the main thesis of this note is very short, but I should like to take the opportunity of replying to a criticism of Good [1966]. Let us first recall the result of that note in precise form.

We refer to 'you', the subject, by the symbol S. Suppose that there are r mutually exclusive and exhaustive hypotheses $H_1, H_2, \ldots, H_r$, and a choice of s acts, or classes of acts, $A_1, A_2, \ldots, A_s$. Let your expected utility of act A_j if hypothesis H_i is true be $U_S(A_j \mid H_i) = u_{ij} (i = 1, 2, \ldots, r; j = 1, 2, \ldots, s)$ and let your initial (prior) probability of H_i be $P_S(H_i) = p_i$. (There will always be some previous evidence, but we take it for granted and omit it from the notation.) You now have the option of performing an experiment, or making an observation, whose possible outcomes are $E_1, E_2, \ldots, E_t$, where $P(E_k \mid H_i) = p_{ik}$ $(i = 1, 2, \ldots, r; k = 1, 2, \ldots, t)$. Here there is no need for a suffix S if we regard the hypotheses as being simple statistical hypotheses, for in this case the conditional probabilities p_{ik} are 'tautological' (having values known by the definitions of the hypotheses and not subject to dispute nor to empirical verification). Suppose further that the cost of the experiment is zero (or negligible), and that the expected utilities u_{ij} and the probabilities p_i have sharp (precise)

---

* This work was partly supported by a grant, No. NIH-R01 GM18770-02, from the Dept. of Health, Education and Welfare (U.S.).

[1] Good [1950], p. 119.

values. Then the theorem proved in Good [1966] is that, before the experiment is performed, the expected utility available to you would not be decreased in the light of the experiment, and would be increased unless the act recommended by the principle of rationality (maximising expected utility) is the same irrespective of the outcome of the experiment. So it is always rational for you to perform a costless experiment.

It has been pointed out to me (by Isaac Levi and Teddy Seidenfeld) that, in the light of my repeated emphasis on the need for non-sharp subjective probabilities and utilities, I should have dealt also with the case where the u_{ij}'s and p_i's are not sharp. The use of 'comparative' utilities and probabilities leads one to the conclusion that there will be situations where there is no rational choice between two options. As Dr Levi said in effect (private communication) the result of an observation might lead to a vaguer state of information than one had before, and so to a state of confusion. I think one way of coping with this criticism is to invoke the 'black box' theory of rationality (see Good [1950], [1952] and [1962]). In this theory, inequality judgments are plugged into the black box, but the calculations within the box are assumed to be performed with sharp values. The box will imply a discernment that we cannot be worse off by performing the experiment, and this result can therefore be regarded as proved if it is assumed that the black box theory is self-consistent. It was proved self-consistent by C. A. B. Smith [1961]. Although this argument seems a little too glib, I cannot see anything wrong with it. At any rate in what follows I shall assume sharp values.

Suppose that some other person, T, who might be a demiurge (T for *Teufel*), has some other set of initial probabilities $P_T(H_i)$ ($i = 1, 2, \ldots, r$) although he agrees with the values of p_{ik} and u_{ij} (for all i, j, and k), and he also knows S's initial probabilities. It might be conjectured that, in T's opinion, S is again better off in expectation if he performs the zero-cost experiment. But the following example shows that this conjecture is false.

To disprove the conjecture we need very little notation because it can be done with a very simple example. We suppose that there are just two hypotheses, either a certain coin is 'fair' or it is double-headed, and that the acts available to S are to bet on one or other of these two hypotheses in a level bet. S must make the bet without examining the coin, but he is given the option of being told the outcome of a single toss of the coin.

It so happens that initially S regards it as just 'odds on' (probability greater than a half) that the coin is fair, whereas T *knows* that the coin is indeed fair. (Or T's estimate of the probability is close to 1.) S might be entirely rational in his opinion, in fact we can imagine that his subjective probabilities are 'credibilities' if such things exist; but T happens to have more information.

Then T knows that with chance one half, the coin will come up heads and that this would cause S rationally to regard the wrong hypothesis as 'odds on'. In this case S will lose his bet. Thus T knows that S has nothing to gain and something to lose if he opts for the experimental information, and therefore T's expectation of S's utility is decreased if S opts for the experiment. T would do S a service by not allowing him to acquire the new information. This would be true even if the experiment consisted of several tosses of the coin, because we have considered a rather extreme situation where T *knows* the truth about the coin. In less extreme situations the result would be true only for experiments involving a small amount of information. When ignorance is bliss, a little learning can be poison according to another man.

I. J. GOOD

Virginia Polytechnic Institute and State University, Blacksburg, Virginia

REFERENCES

Good, I. J. [1950]: *Probability and the Weighing of Evidence.*

Good, I. J. [1952]: 'Rational decisions', *Journal of the Royal Statistical Society B*, **14**, pp. 107–14.

Good, I. J. [1962]: 'Subjective probability as the measure of a non-measurable set', in *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress* (Stanford), pp. 319–29.

Good, I. J. [1966]: 'On the principle of total evidence', *British Journal for the Philosophy of Science*, **17**, pp. 319–22.

Smith, C. A. B. [1961]: 'Consistency in statistical inference and decision', *Journal of the Royal Statistical Society B*, **23**, pp. 1–25.

Brit. J. Phil. Sci. 25 (1974), 343–397 *Printed in Great Britain*

Review Articles

OTTO NEURATH[1]

This is the first volume in what promises to be a valuable collection (expected to grow eventually to some thirty volumes) of works relating directly or indirectly to the Vienna Circle.

With an eye to the volumes to come, I will begin with some criticism of the editing of the first volume. There is no editorial introduction to it, only a general introduction to the series. Nor are the separate items introduced (other than by a foot-note reference to a numbered item in the bibliography at the back of the volume). And there is no subject index. After a biographical chapter (this is very good; I will revert to it) the reader comes first to some Chinese-style fables signed 'La-Se-Fe'; a 'Note by the Author' says that they have been translated by Peter Zirngibel (Dresden) from *Convent of the Seven Blessings*, Peking, 1904. What are these doing here? Only when he is half-way through this chapter is the puzzled reader told that 'La-Se-Fe' and 'Peter Zirngibel' were among the pseudonyms used by Neurath. Why not tell him straight away? And why include these philosophically unrewarding bits of chinoiserie anyway?

This brings me to a larger complaint. Since there is no editorial introduction there is no disclosure of the editors' principles of selection; and one wonders what these were. The interests of Neurath that dominate this compilation are: Marxism and Isotypes. For the opening volume in a 'Vienna Circle Collection', there is surprisingly little of Neurath's Vienna Circle work. None of his contributions to *Erkenntnis* is included. There is nothing on the idea of an encyclopaedia of unified science (though the bibliography includes eight items on this subject). Neither 'Protokollsätze', nor 'Pseudorationalismus der Falsifikation', nor 'Le développement du Cercle de Vienne . . .' is translated here. And most of Neurath's celebrated sayings are absent: there is no 'universal slang', no 'Otto's protocol at 3:17 o'clock . . .', no sailors re-timbering their vessel on the open sea. Indeed, the only piece included here that is directly related to the work of the Vienna Circle is the famous manifesto, 'Wissenschaftliche Weltaufassung: der Wiener Kreis'. According to Philipp Frank (whose recollections of Neurath might have been included in the biographical chapter), it was Neurath who invented the name 'Wiener Kreis'—he thought it would give 'The Scientific Conception of the World' cheerful associations with the Viennese waltz and other pleasant things.[2]

Incidentally, the production of this manifesto required Neurath's political skill. Schlick abhorred the idea that his seminar-group should turn into a philosophical sect with a corporate doctrine, and so could not be expected to sympathise with the idea of a group manifesto. So Neurath waited until Schlick was safely away in America. Then he drafted a manifesto which cast Schlick as founder and centre of the Circle. It was vetted by Carnap and Hahn, who signed it with Neurath. According to Heinrich Neider (p. 49), Schlick 'was not very enthusiastic about its formulations, especially those of a political or worldview character' (it connects 'modern empiricism' with the 'socialist attitudes' of 'the masses', p. 317). But Schlick could hardly repudiate it: for it was dedicated to him and the preface stated that 'the present pamphlet on the Vienna Circle . . . is to be handed to Schlick in October 1929 when he returns from his visiting professorship at Stanford University, California, as a token of gratitude and joy at his remaining in Vienna' (p. 299).

---

[1] Review of Neurath, O. [1973]: *Empiricism and Sociology*. Edited by Marie Neurath and Robert S. Cohen. Dordrecht-Boston: D. Reidel. Vienna Circle Collection, 1. Pp. xx+473.

[2] Philipp Frank [1949], p. 48.

The opening chapter consists of a highly successful biographical pastiche composed of fragments written by Neurath himself, his father, his widow (who is co-editor), his son, and by Carnap, Kraft, Morris, Popper, and others. One gets a vivid impression of a man of enormous vitality and ebullience, an omnivorous reader, a tireless and overpowering conversationalist, a forceful and effective organiser with a sharp eye for detail, and a revolutionary socialist. After 1918 'the revolution attracted him with magical power' (Ernst Lakenbacher, p. 14). He had an opportunity to put socialism into practice when he became President of the Central Planning Office in Munich under the Socialist government of Bavaria, in March 1919. It was a turbulent period. The prime minister had just been murdered. Neurath stayed on when a 'soviet republic' was proclaimed. Civil war broke out. By mid-May Neurath was under arrest, and was afterwards imprisoned, briefly, for 'treasonable' activity.

In the 1920s he was busy with, among other things, the Social and Economic Museum of Vienna and, of course, the Vienna Circle. Schlick and Carnap may have been the Circle's intellectual leaders, but it was largely Neurath's dominant personality and organising drive which transformed the Circle from an informal discussion club into a powerful philosophical cadre with its own journal, congresses, *etc.*

When revolution and counter-revolution broke out in Vienna in February 1934, Neurath was in Moscow; it was too dangerous for him to return, so he moved to the Hague where, anticipating that Vienna would become impossible, he had already founded a Visual Education Centre. (This last part of his story is beautifully told by his widow.) In May 1940 they managed to escape in an over-crowded boat to England. ('I speak broken English fluently', he had cheerfully announced in 1936.) They were released from internment in February 1941, and moved to Oxford, where they founded an Isotype Institute. He lectured on logical empiricism at All Souls. On 22 December 1945, he was in his study with Marie Neurath. He laughed at something; she looked up—and found him dead.

He possessed the kind of charisma that gets on some people's nerves. Schlick declared that he could not bear Neurath's loud voice, according to Heinrich Neider (p. 48), who adds: 'These two men were opposites not only in their thinking but also in their personalities. Schlick, the soft-spoken tall man, supposed to be of aristocratic parentage, lived in the style of the grand bourgeoisie—he used to ride on horseback before his lectures—loved to travel, gave musical parties and dinners in the style of cultured rich people' (p. 47), whereas Neurath was loud, bohemian, and boisterous.

So much for the man; now for such of his writings as are included here.* They range from a 1912 paper on 'The Problem of the Pleasure Maximum' (to which I will revert) to an unfinished 1945 paper on visual education, with papers on optics and on economic planning in between. And the quality is variable. There are some striking things. There are also passages which seem, at least to this politically unsympathetic reviewer, to lapse into Marxist propaganda (*e.g.* pp. 290, 297, 363), or into near-banality (*e.g.* 'Who needs splinters must either wait for the splitting of wood to happen or do it himself', p. 406). Also, Neurath had a habit of taking the stuffing out of bold-seeming statements by prefacing them with 'to a certain degree' or 'under certain circumstances' or 'in a certain sense' or 'within a certain domain', without indicating *what* degree, circumstances, sense, or domain.

Is there a thread running through this variegated material which, if followed, might give some coherence and pattern to a discussion of it? I think there is. I have noticed a persistent and striking parallelism, or rather linkage, between Neurath's philosophical ideas and his political ideas. (We touched on an example of this in the 'Wiener Kreis' manifesto.) I will consider four such philosophico-political links.

(1) To begin with, there was for Neurath a close connection between his physicalism (which he regarded as an up-dated version of materialism) and his Marxism. His argument for the tie-up seems to go something like this: physicalism, and the demand for a unified science, requires a physicalist *sociology*; only Marxism provides such a sociology; so physicalism impels us towards Marxism. Once one recognises that 'physicalism succeeds materialism' (p. 359) one will 'no longer hesitate to speak of a "materialist unified science" and a "materialist sociology" ' (*loc. cit.*); and '*the living contemporary form of such sociology is Marxism.* What sociology on a materialist foundation essentially is like can best be clarified by analysing the materialist conception of history . . .' (p. 363, his italics).

The anti-metaphysics of the Vienna Circle was obviously a kind of *Ecrasez l'infâme!* But what particular *infâme* were they mainly concerned to crush? With the disadvantage of hindsight I had supposed that the threat of Naziism had given them a horror of Germanic metaphysics and especially of Hegelianism. But this does not fit Neurath. He quotes Hegel at length and without disapproval on the historic destiny of the German people (pp. 336–7), and comments approvingly on Hegel's recognition of 'jumplike changes' in history. His appraisal of Hegel was more Marxist than positivist: 'Hegel's doctrine was mostly used to support the regime. . . . But from Hegel's exposition one can draw quite revolutionary reflections too' (p. 348). Hegel's 'gigantic theodicy', far from being dismissed as metaphysical mumbo-jumbo, seems rather to have been regarded by Neurath as a precursor of his own idea of the unity of science: Hegel's 'philosophy of history inspired a comprehensive description of the whole course of society'; this meant that Hegel 'must endeavour to give a uniform account of all factual sciences. This was largely the case' (p. 336). Of course, this theodicy, with its demonstration of a divine dialectical order, needed to be de-theologised:

---

* [*Editors' note*]: Professor R. S. Cohen has prepared a Note on Otto Neurath's writings in English and in English translation. We hope to publish this in a later issue of the *Journal.*

> 'the true [i.e. Marxist] scholars tried scientifically to preserve the order thus demonstrated, and to eliminate God' (p. 338).

If not Hegel, then who or what was the *infâme*? There is in this collection a long (and rather rambling and disappointing) attack on Spengler; but that was published in 1921, and does not employ positivist weapons. For Neurath, it seems, catholicism and bourgeois ideology were the *infâme*, to be crushed by logical empiricism and physicalist sociology: 'the revolutionary masses of workers . . . become more vigorous through anti-metaphysical physicalist sociology, and above all the fight against metaphysics and theology means the destruction of bourgeois ideology' (p. 356).

But how could physicalist sociology, which 'knows only of such behaviour of men that one can observe and "photograph" scientifically' (p. 361), have such an invigorating effect on workers? The answer, so far as I can discern it, seems to be this. Physicalist sociology, though descriptive, is not *merely* descriptive. There is no such thing as sociological neutrality (p. 417). Ethics (like metaphysics) has been exploded, but the better remnants have somehow got absorbed into the new sociology. That, at least, is how I construe this passage:

> A person of scientific attitude feels bound to organisations of men and behaves accordingly. *In place of 'ethics', we have systematic sociology*, in place of 'ethical' behaviour, behaviour within a group, determined socially. Certain endeavours to maintain a kind of 'scientifically' acceptable 'ethics' will continue for a while, until an incorporation into common action, and *a scientific sociology as basis for* it and for *all collective behaviour*, will have come into their own (p. 328, my italics).

(2) Another link between Neurath's physicalism and his socialism is his physicalistic conception of economic planning. So far as I know, he did not draw an explicit analogy between money and metaphysics; but he seems to have regarded money, rather as he regarded metaphysics, as a centuries-old sham which needs to be exposed so that it will no longer bewitch and befuddle people: 'Once the nature of money will have been fully recognised, the scales will drop from everybody's eyes, and the development of centuries will look like one great error' (p. 145). Also, from a planner's point of view, money is slippery stuff: 'It is the essence of money that it cannot be directed'; so, 'in a socialised economy, money no longer is a driving force. . . . Money could remain at best as a token . . .' (*loc. cit.*) Banks might be allowed during the transition period; but it seems that people would be unwise to pay money into them, since 'withdrawals should not be allowed' (p. 148). And after the transition period? 'What then will replace net profits in a socialised economy?' (p. 146). Answer: 'We shall simply have to determine production and consumption . . . by direct consideration of the various possibilities' (p. 147). Rather as Leibniz supposed God to select the best out of all possible worlds, so Neurath expected the Central Economic Office to consider in full detail all the possible ways in which the physical resources of the country might be used ('one can prefer one plan to another only on the basis of a total estimate', p. 264), and then choose. 'Those who shy away from this should keep their hands off socialism' (p. 147). This was written in 1919.

The idea of physicalist economic planning which Neurath was advocating in

1919 was not presented as a distant utopian ideal, but as something to be put into effect forthwith (with Neurath himself playing a role). 'At present', he wrote, 'we live in a period of conscious shaping of life' (p. 150). And:

> Successful socialisation is possible only of the whole and from above. If one wants to socialise at all it should be done at once and quickly . . . (*loc. cit.*)[1]

What considerations should govern the choice between rival plans? Presumably the happiness of people will be the main consideration. The problem of measuring happiness occupied Neurath, off and on, for at least thirty years. It is a difficult problem, especially when approached from a physicalist standpoint that allows only 'photographable' behaviour to be taken into account. Neurath's attitude to it oscillated. Sometimes he came close to conceding its insolubility; at other times he made programmatic assertions that it will be solved, though without vouchsafing how. An early (1912) discussion of the problem ended on a defeatist note:

> Perhaps struggle will decide which view about the best order of life shall be victorious; perhaps preference will be given to one order out of those in question, and the choice may be made with the help of an inadequate metaphysical theory or in some other way; tossing coins would be much more honest (p. 122).

In 1931 he took an optimistic line:

> One can now begin to compare the standards of living of different men on the basis of objectively ostensible characteristics. . . .
>
> If one wants to treat these problems more accurately, one must set up a scale of living standards, which would be comparable to the hardness scale in mineralogy (p. 401).

In 1942 he claimed that 'happiness conditions' can be represented by Isotype diagrams (which I will discuss later), representing visually the amounts of 'food, shelter, accidents, bullying, civil liberty' etc. associated with a certain walk of life (p. 424). These 'silhouettes' of 'happiness conditions' may then be compared with one another.

But if one silhouette excels another on some counts and is excelled by the other on other counts, is there any way of summing their respective values to discover which on balance is best? Neurath conceded that there is not:

> We have no unit for measuring the various items and for thus computing them and forming one characteristic number. No 'value reckoning' helps us here and moreover 'mathematical economics' would lead us astray (p. 425).

Thus even if a Central Economic Council, composed of ' "Brain trusts" of first-class scientists' (p. 426), *could* survey all physical possibilities, it could, on Neurath's own admission, hardly make a non-arbitrary choice between them.

---

[1] By 1928 this had become a rather distant ideal: '*Perhaps the time will come* when occupations and men will be considered as a connected whole, which of course presupposes that *the total process of production and distribution is ruled according to a plan*' (p. 282, my italics).

In some ways Neurath was more in the tradition of Saint-Simon (who does not figure in this volume) than of Marx. Here is a description by Hayek of Saint-Simon's *Introduction aux travaux scientifiques du xix*^{e} *siècle* (1807–8):

> Yet, with all its defects, the work is a remarkable document. It combines, for the first time, nearly all the characteristics of the modern scientistic organiser. The enthusiasm for physicism (it is now called physicalism) and the use of 'physical language', the attempt to 'unify science' and to make it the basis of morals, the contempt for all 'theological', that is anthropomorphic, reasoning, the desire to organise the work of others, particularly by editing a great encyclopedia, and the wish to plan life in general on scientific lines are all present.[1]

(3) I now turn to another philosophico-political link, this time between Neurath's idea of a unified physicalist language and his proletarian socialism. And I will go on to suggest that each of these connected ideas was extended—the former into the pictorial language of Isotype, the latter into the idea of international socialism—and that there was again a link between their respective extensions.

However one appraises Neurath's thesis of the reducibility of all scientific theories and concepts to a physicalist observation language, together with its corollary that '*Physicalism knows no "depth", everything is on the "surface"*' (p. 326, Neurath's italics), one must surely sympathise with the motive behind his reductionism. This was a concern for what he called 'the humanisation of knowledge' (p. 232). In a striking passage he declared that the Middle Ages will end as 'soon as all men can participate in a common culture and the canyon between educated and uneducated people has disappeared' (p. 224). Statements expressed in a physicalistic observation language can be readily understood by ordinary people. But as they stand, the theories of contemporary science are accessible only to a small intellectual elite. Thus if those theories could be *reduced* to a 'common neutral and democratic language' (p. 232) they would be humanised and democratised. Neurath expressed this picturesquely and optimistically when he wrote:

> One day Einstein's relativity theory will be taught to everybody as simply as Kepler's theory is today. It seemed very strange and difficult to his contemporaries, but today it is demonstrated by means of oranges, knitting needles and candles. (*loc. cit.*)

That science should be available to all, and not just to a few (mainly bourgeois) intellectuals, is a noble idea; at the very least, we can surely agree that there should be no class divisions in the republic of scientific knowledge. Unfortunately, this was not exactly Neurath's view. What he wanted was not just the democratisation but the *proletarianisation* of science. For him, the struggle of science against metaphysics was bound up with the struggle of the proletariat against the bourgeoisie:

> It is precisely the proletariat that is the bearer of science without metaphysics (p. 297).

---

[1] Hayek [1952], p. 123.

At another place, Neurath mentions the fear of 'some representatives of the proletarian struggle' that science and culture may be damaged by revolution; he comments:

> And so, all the more significant it is that closer investigation can show that scientific thinking is safeguarded best by the proletariat which can use it to best advantage, whereas among its adversaries an unscientific attitude often strengthens the front-line (p. 252).

(The editors report (pp. 82–3) that Neurath exasperated Schlick with his frequent interjections of 'metaphysics!' during the Circle's discussion of Wittgenstein's *Tractatus*. One wonders whether this rhetorical passage does not invite a similar interjection. Just what does 'closer investigation can show . . .' mean? Whatever it means, the passage does not meet the serious question it purports to answer.)

(4) I now turn to the link between the internationalist aspect of Neurath's socialism and his development of the idea of a physicalist language into a new kind of pictorial language (Isotype: the name was invented by Marie Neurath) that could be internationally understood.

I will not dwell much on the former. Briefly, Neurath believed that world peace requires socialist revolutions in all capitalist countries (see, *e.g.* pp. 270–1) resulting, eventually, in a 'socialist world empire' (p. 289) organised according to 'a world economic plan' (p. 273). (He was not a 'minute philosopher'.)

In Neurath's Isotype language, information and statistics are conveyed by simple, stylised pictorial symbols. For example, the death-rate of a country at different periods might be represented by rows of coffins, each coffin standing for so many deaths a year. There were various ideas behind this language. One was that the human mind has always been more receptive of images than of words, and that this tendency is being aggravated in the modern world:

> Modern man is conditioned by the cinema and a wealth of illustrations . . . If one wants to spread knowledge, one should use means similar to modern advertisements (p. 214).

(This was written in 1925, in connection with the Social and Economic Museum in Vienna.)

Also Isotype was a kind of fulfilment of Neurath's physicalist empiricism. Classical empiricism had conceived all knowledge as growing out of what were, in the end, *private* images. (Something of this idea was retained in the 'methodological solipsism' of Carnap's *Aufbau*; Neurath, and Popper, persuaded him to abandon it.) In Isotype, knowledge is encapsulated in *physical* images.

But it seems that the main motive for the development of Isotype was Neurath's belief that it would overcome language barriers between peoples. Carnap was an enthusiastic Esperantist.[1] But Esperanto has to be learnt. Isotype was intended to be immediately understandable by anyone in any part of the world. (Actually, quite a few of the illustrations of it given in this volume have had to be accompanied by some verbal explanation. I do not know what a Bantu would make of the representation on p. 242 of the spreading of tuberculosis. Also,

---

[1] See Carnap [1963], p. 69.

Isotype is unlike Esperanto in being an essentially didactic language: information and instructions can be imparted, but the recipient cannot answer back.)

Neurath's motto, in this connection, was '*Words divide, pictures unite*' (p. 217).

> Visual education leads to internationalisation much more than word education does (p. 247).
>
> Besides being more international, visual education from the beginning enforces a more human approach than education through words. It can only deal with things which are within the grasp of everyone's understanding.... I wish to promote visual education as an element of human brotherhood (p. 248).

Isotype soon caught on in some places. Already by 1933 Neurath could report:

> In the Soviet Union, the Council of People's Commissars has issued a special decree to the effect that 'Dr Neurath's method of graphic representation of statistics is to be applied in all schools, trade unions, public and co-operative organisations' (p. 222).

(A happy personal touch of Neurath's was to sign his letters with Isotype-like drawings of an elephant; see p. 65.)

In conclusion, I want to look into a question which obviously agitated Neurath, namely the question of the predictive power of the 'new' empirical and physicalist sociology.

Here, two things should be borne in mind. First, by Neurath's own yardstick, this sociology was either a genuine part of unified science, or it was nothing; and a putative scientific statement has 'significance only in so far as it can be tested empirically' (p. 313). Second, Neurath imputed to this sociology a role of truly colossal importance: it was to provide the knowledge on the basis of which a 'total transformation' (p. 261) of the whole structure of national and international life was to be effected; and after the transformation a huge amount of sociological knowledge would continue to be needed for the planning of the 'socialist world empire'. Clearly, it would need to be very good sociology; which meant, for a Vienna Circle empiricist, that it should be predictively powerful, and that its predictions should be pretty successful. Neurath endorsed this: '*Prediction* of the coming social structure and of the functioning of a given social structure is then *at the centre* of a planned way of life' (p. 403, my italics).

The concluding section of his *Empirical Sociology* (1931), entitled 'Sociological Prognosis', is concerned with just this question. It contains some confident-sounding pronouncements:

> Sociology improves its predictions, and on the whole it can predict with growing success (p. 407).
>
> . . . *scientific sociology* . . . *tells the masses what the future will bring them* (p. 409, his italics).

But I get the impression that this was whistling in the dark. The first of these quotations concludes a long discussion of the *limitations* of sociological prediction: Neurath stressed that inventions cannot be predicted (p. 405) and that sociological predictions have a disturbing influence on their subject-matter; he insisted that 'the predictions we make must remain indefinite on many points', and that 'all our predictions can assume several forms' (p. 407).

More telling than these admitted limitations and qualifications, however, is the fact that in this section Neurath does not produce one genuine specimen of a sociological prediction made in advance and logically capable of being refuted by subsequent developments. Marx ventured predictions of this kind. As we saw, Neurath considered Marxism 'the most modern sociology on a materialist basis (which in the end means physicalism, the most modern form of materialism)' (p. 410). So one might have expected Neurath, in a section explicitly devoted to sociological prognosis, to look carefully at the empirical success or failure of Marx's main predictions. He leaves this question untouched. What he does adduce, in support of the prognostic power of his sociology, are various pseudo-predictions: that is, 'predictions' which, Neurath now claims with hindsight at time t_2, could have been made at t_0 about a development now known to have occurred at t_1. For instance:

> By carefully examining the situation and structure of classes in Russia around 1900, one can predict the mechanism of the revolution which was then preparing itself (p. 410).

Again:

> If [the sociologist] wants to deduce the fate of Europe in 1850 from its fate in 1800, he will above all try to describe the mass situations and to establish relations which allow prediction . . . (p. 370).

At one place (p. 372) he declares: 'That economic crises occur, that rationalism develops, can be structurally deduced . . .', but without saying *what* sociological premisses these things can be deduced from. At another place he lays down what sort of thing should go into the predictive premisses, but without mentioning *what* predictions would be deducible therefrom:

> For instance, 'historical materialism' would seek to base its predictions about the U.S.S.R., where the swiftest changes are now [1931] taking place, on the present class structure. It would note that above all it is industrial and agricultural workers that are decisive, that the 'kulaks' are being pushed back, that the groups of 'new bourgeoisie' will hardly play an important role. But it is already difficult to predict in detail how far this socialist total organisation will succeed or fail. . . . Nevertheless, if one succeeds in making some statements, it will be on the basis of Marxist analyses as sketched here . . . (p. 411).

The suspicion begins to dawn, at least on this sceptical reviewer, that Neurath's 'Empirical Sociology' could yield *no* predictions and was therefore without scientific content; that the great system which was to guide the proletariat in bringing about a 'total transformation' should have been regarded, from Neurath's strict empiricist point of view, as on a par with the despised systems of metaphysics and theology with which the bourgeoisie propped up *their* political aspirations.

There is even some evidence in the present volume that Neurath shared this suspicion. His son Paul reports that his father persistently tried to dissuade him from becoming a sociologist: 'He did not think that this was a "field" with a "subject matter" ' (p. 39).

J. W. N. WATKINS
*The London School of Economics
and Political Science*

REFERENCES

Carnap, R. [1963]: 'Autobiography' in P. A. Schilpp (*ed.*): *The Philosophy of Rudolf Carnap*, The Library of Living Philosophers (Open Court).

Frank, Philipp [1949]: *Modern Science and its Philosophy* (page reference to the 1961 Collier Books edition).

Hayek, F. A. [1952]: *The Counter-Revolution of Science* (Free Press of Glencoe).

ORTHODOXY IN QUANTUM MECHANICS[1]

The formalism we now know as non-relativistic quantum mechanics (*QM*) was developed in a great flowering of theoretical activity over a very short space of time (1925–7). Different points of departure, associated with the ideas of Heisenberg and Schrödinger, were rapidly united in the transformation theory of Jordan and Dirac, while the mathematical shortcomings of Dirac's work were surmounted in a theory of amazing power and beauty by the mathematician von Neumann, the so-called Hilbert space formulation of *QM*.

But it must be remembered that these theories were developed, not as abstract mathematical exercises, but in direct response to the extraordinary experimental discoveries which had built up during the preceding quarter of a century to challenge the concepts and methods of classical physics. From its inception then the question of the interpretation of *QM* was well to the fore, but in marked contrast to the rapidity with which the formalism came to be accepted by physicists, mainly due to its quite extraordinary explanatory power over the whole range of atomic and molecular physics as those subjects were then known, debates and arguments about what the formalism might be interpreted to mean, and what were the philosophical implications of the proposed interpretations, have continued over a period of nearly fifty years, and remain at the present time as an outstanding challenge to philosophers of science.

In his book *The Logical Analysis of Quantum Mechanics* Professor Scheibe has set himself the task of presenting a careful and exact analysis of some of the outstanding problems from the so-called Copenhagen, or orthodox, point of view. The name 'Copenhagen' derives from the fact that the major contribution to the development of the orthodox interpretation was made by Niels Bohr, referred to by Heisenberg as the source of the 'Kopenhagener Geist der Quantentheorie'. Undoubtedly Bohr's approach to the interpretation of *QM* is philosophically the most interesting and penetrating of any of the quantum theorists, but his views were not held, at any rate without some modification, by some who would undoubtedly have regarded themselves as his disciples. Both Heisenberg, and to an even more marked extent von Neumann, departed from the Bohrian position, and the mainstream orthodox exposition is usually a blend of different strands deriving from a variety of sources. Anyone endeavouring to present an exhaustive account of *QM* must endeavour to draw these strands apart and exhibit *QM*, not as a single formulation, but as a possibility of different interpretations which may then be critically compared and assessed. Scheibe does not actually attempt such an undertaking in its entirety. He is more concerned with presenting a number of ways of formulating some selected problems of *QM* within orthodoxy on the grounds that this may prepare the reader for taking part in the debate between orthodoxy and its critics, or indeed between the different varieties of orthodoxy. As Scheibe puts it '. . . the immediate question is: *what* is being debated?'. This limited objective has useful and not so useful consequences. The useful consequence is that there is space for very detailed discussion of what *is* discussed. The disadvantage is that one is in no position, as a result of reading this book, to defend orthodoxy (or to reject it) since the criticisms levelled against it are only mentioned in passing. Perhaps a sequel setting out the critical positions in as much detail as the orthodox theory will be usefully forthcoming from Professor Scheibe's pen.

---

[1] Review of E. Scheibe [1973]: *The Logical Analysis of Quantum Mechanics*, Oxford: Pergamon Press, £6.50. Pp. viii + 204.

Turning to the scheme of the book under discussion, Scheibe begins with an interesting chapter setting forth the main lines of Bohr's thought as expressed in the many essays he wrote, which usually have the deceptive appearance of popular or semi-popular expositions, but which reveal the central fact that Bohr saw in *QM* the working out of a general philosophy of nature. In fact it is clear that Bohr was convinced that in his idea of complementarity lay the key to epistemological puzzles arising in many aspects of human experience. Scheibe confines himself to Bohr's remarks in connection with physics, and his treatment here is very complete and lucid, bringing out the essential ideas in a way which invites favourable comparison with the recent critiques of Bohr's philosophy by Feyerabend [1968], [1969] and by Hooker [1972]. The most remarkable thing about Bohr's contribution to the foundations of *QM* is that it proceeds with the minimum of formal development. Bohr places the greatest emphasis on drawing out the correct conceptual framework within which to analyse the quantal description of atomic phenomena. This must precede the mathematical formulation, which should simply express the conceptual analysis which Bohr believed was suggested, or even necessitated, by the experimental situation with regard to the complementarity of the wave and particle pictures of matter and of radiation. Scheibe discusses the difficult question of the rôle played in Bohr's thinking by his analysis of idealised or gedanken experiments. He concludes, I think correctly, that these were intended by Bohr to exemplify his general position, and were in no sense used to go beyond or extend the ideas originated in his more general discussions. Their importance lies mainly in demonstrating the self-consistency of Bohr's ideas when these were applied to measurements on the apparatus used in some primary measurement, phenomena wherein Einstein and other critics of Bohr sought to demonstrate some kind of logical contradiction within orthodox *QM*.

In the second chapter of his book Scheibe takes up the task of presenting the mathematical formulation of *QM*, basing the work mainly on the ideas of von Neumann. He carefully distinguishes four possible different methods for developing the theory, posing first of all the question of whether states should (ontic formulation) or should not be assigned to individual quantum objects, and secondly correlating these two possibilities with the interpretation of probability statements either as expressing states of knowledge about individual objects (epistemic interpretation) or as specifying relative frequencies in statistical ensembles of individual objects (statistical interpretation). His analysis of the concept of probability in *QM* is somewhat superficial, and as usual he makes no attempt to assess the relative merits of the different approaches he is at pains to distinguish. He simply takes over in an uncritical manner, effectively, the subjective and frequency interpretations of probability. In regard to the ontic formulation he recognises a third interpretation which appears to have close affinities with Popper's propensity theory, regarding probability as an objective property of an individual object in relation to a repeatable experimental arrangement for observing it. But again Scheibe's account is disappointing and fails annoyingly to discuss the philosophical arguments for or against his ontic formulation. The question of how classical probability theory (as represented for example by the Kolmogorov axioms) may need to be modified in the quantum-mechanical context is not raised explicitly. The work of Varadarajan ([1962]) and Prugovečhi ([1967]) is referred to, the latter very briefly, later in the book, but the important contribution of Fine ([1968]) to this discussion is not mentioned.

Scheibe takes some pains to express *QM* as a generalisation of an indeterministic formulation of classical mechanics (following essentially ideas presented by von Neumann many years ago). While this is an interesting topic in its own right, the relevance to characteristic features of *QM* of the admitted inadequacy in idealised assumptions about exact measurements of continuous variables is probably minimal. Such points are of great interest when one is discussing difficulties associated with theoretical extensions of *QM*, for example in relativistic quantum field theory or elementary particle physics generally, but in non-relativistic *QM* the essential introduction of such ideas, which von Neumann, incidentally, found a pleasing feature of his formulation as against Dirac's, is something of a side issue.

A brief chapter follows dealing with the formulation of *QM* in terms of properties rather than quantities (essentially one asks, instead of what values does some observable have, the related question does this observable have such and such a value: the answer to the first question is a number, or perhaps a set of numbers, the answer to the second is yes or no). Scheibe then turns to the problem of determinacy versus indeterminacy in the formulation of a theory with particular reference to *QM*. His ideas here are not particularly novel. *QM* is deterministic in so far as the state description varies in a precise way with time in accordance with the time-dependent Schrödinger equation, so long as no measurements intervene; it is indeterministic in so far as the specification of state in *QM* is a probabilistic one. When he comes to discuss the significance of the Heisenberg uncertainty relations Scheibe makes some useful distinctions as to what one might want to mean by the uncertainties involved. He proceeds to give a clear analysis of the usual one and two slit experiments which bring out so many of the puzzles of *QM*. He refers briefly to the Beck and Nüssenzweig analysis of diffraction by a slit but does not emphasise the point, not always maintained in the literature, that their treatment cannot properly be used to furnish a counterexample to Heisenberg's uncertainty principle in the case of a narrow slit.

At this point in his book Scheibe interpolates a short chapter on what he calls the von Neumann programme. Essentially Scheibe distinguishes two ways of presenting a theory in physics. Firstly we may select a class of physical concepts and a class of mathematical structures and represent the former in terms of the latter. The disadvantage of this procedure according to Scheibe is that the mathematical theory in general contains a lot of 'surplus' structure having no physical interpretation. Scheibe therefore endorses the programme, initiated in *QM* by von Neumann, of attempting a direct axiomatisation of the physical concepts. Scheibe's enthusiasm for such an approach, which he regards as much more profound, appears to be engendered by an underlying commitment to an uncompromisingly realist view of theories, and the correctness of his appraisal of the rôle of mathematics in theoretical physics is certainly arguable. The surplus structure of the mathematical theory may provide the means most suitable for modifying or adjusting the physical theory—the example of analytic *S*-matrix theory in modern elementary particle physics is an excellent example of such a possibility.

In the final two chapters of the book Scheibe addresses himself to two key topics in the understanding of orthodox *QM*: the reduction of states and the vexed question of completeness. By the reduction of states is meant the change in the state of a quantum-mechanical system consequent on measurement. In orthodox *QM* the process of measurement is supposed to convert an initially pure state into a final mixture which would correspond to a transition (in Heisenberg's phraseology) from a potentiality for various possibilities to the actualisation of these possibilities, from an objective probability of tendency to a subjective probability of incomplete knowledge (as to which final state has in fact been actualised). But such a transition (pure state to mixture) is formally inconsistent with the time-dependent Schrödinger equation. How is this contradiction to be resolved? There are two main lines of approach: (*1*) we accept that measurement is quite a different process from ordinary dynamical interaction and the law governing the change of state is accordingly also different, (*2*) we try to effect some sort of reconciliation between the two ways in which states can alter. This second alternative was von Neumann's way and is the objective of a wide range of literature dealing with the theory of measurement in the context of orthodox *QM*. Scheibe follows von Neumann closely in his general approach and claims to prove that weak reduction (*i.e.* in so far as measurements of observables which correlate object and apparatus are ignored, the required mixture is obtained) is feasible, whereas strong reduction (the final state is a mixture even with respect to correlation measurements) is not possible. This was von Neumann's own conclusion and Scheibe's re-writing of his ideas does not seem to add anything, except a greatly more complicated presentation. My main criticism of Scheibe here is that he quite overlooks the important work of Wigner [1952] and Araki and Yanase [1960] which showed that the von Neumann formulation of the measurement process was inconsistent with the measurement of quantities referring to the object which failed to commute with additive (*i.e.* with respect to the Hilbert spaces of object and apparatus) quantities which were conserved during the measurement. A correct formulation of the measurement problem within orthodox *QM* must allow for the resulting non-ideality of measurement (using d'Espagnat's terminology). The possibility that the apparatus is initially not in a pure state but a mixture should also be allowed for. (This point *is* dealt with by Scheibe.) Such an analysis (which incidentally fails to solve the measurement paradox) has been supplied by d'Espagnat [1966] (see also Earman and Shimony [1968], Stein and Shimony [1971]). But none of this well known work is referred to by Scheibe. I regard this omission as the most serious defect in a book devoted to such a detailed and careful account of the orthodox position as Scheibe claims to provide.

Having failed to justify the strong reduction of states within the von Neumann analysis, Scheibe gives no attention to the two main approaches to solving this problem within a uniform quantum—mechanical description of nature, *viz.* to utilise the classical nature of the measuring apparatus on the one hand, or the irreversible features of the measurement process on the other. The relevance of Bohr's ideas, expounded so clearly by Scheibe in his opening chapter, to these modern approaches could have been usefully discussed. The reviewer's own opinion is that Bohr's views are much closer to the 'classical apparatus' programme as expounded most clearly for example in Jauch, Wigner and Yanase [1967] than to the irreversibility approach, whose leading protagonist has been Bohr's old disciple, the late Professor Rosenfeld (lending the weight of his authority to the Italian school and more recently the Brussels school of quantum statistical mechanics). My reading of Bohr would be to support the thesis that irreversibility may complete a phenomenon but plays an essentially secondary role in objectifying it. Bohr himself, of course, had no patience with the view that strong reduction was a feature of human consciousness, as believed by von Neumann himself, and advocated in more recent times by Wigner. But Bohr's true attitude to the measurement problem is hard to disentangle, oscillating between a metaphysical position regarding measurement as an irreducible constituent of the epistemological framework, and a practical down-to-earth view that measurement interactions are ordinary quantum mechanical processes. This equivocation is of course in keeping with Bohr's constant emphasis on the dual role of apparatus as observer (classical description) and potential object of observation (quantum-mechanical description).

Finally there is the question of completeness. The point here is whether in principle or practice the limited probabilistic description of states in *QM* could or indeed, as claimed by Einstein in the famous Einstein-Rosen-Podolsky paradox, should, be augmented by a more complete description. The first possibility leads to a discussion of hidden variables in *QM* and the various 'proofs' for the impossibility of introducing them. The validity of such proofs depends on what one is prepared to admit as a legitimate hidden-variable theory (Kochen and Specker [1967] demonstrate for example that hidden variables of a sort are always possible). Scheibe does little more than review the literature on this problem with particular emphasis on the rather unfamiliar work of Kamber.

Next Scheibe discusses the *ERP* paradox. He stresses that Einstein did not here try to prove that *QM* was inconsistent within itself, but that elements of reality existed which had no counterpart in the theory, hence one should expect it to be augmented to accommodate them. Scheibe re-writes the argument by first extending *QM* to include essentially definitions of what it means for a quantity to be real in a state and what it means for knowledge of a state to be complete. Introducing the famous *ERP* criterion for an element of reality (certain knowledge can be acquired without any physical disturbance of the system), Scheibe demonstrates a strict logical contradiction. As a variation he discusses the more usual version of the paradox in terms of acquiring information about one physical system by observing another with which it has interacted in the past. Scheibe concludes an extended and very precise analysis of all aspects of the paradox by referring to Bohr's solution, *viz.* it all depends what one means by the expression 'without in any way disturbing a system'. The conditions for defining the value of a quantity referring to the first system are inextricably involved with the observations we make on the second system, and the non-separability of quantum objects is a fact one must learn to live with, however it may offend a classically inspired intuition. There is no doubt that the *ERP* argument caused a radical deepening of Bohr's philosophical insight, and as Scheibe rightly stresses, Bohr's concept of an indivisible 'phenomenon' as the primary object of quantum-mechanical discussion was first really crystallised in his reply to *ERP*.

In summary then Scheibe's book continues the tradition of Pergamon's International Series of Monographs in Natural Philosophy (of which this work is *Volume 56*) in providing detailed and authoritative accounts of important areas of physics. The major criticism is that the book has no unifying theme, it is really a selection of very detailed discussions on disjoint problems, rather as if the author were preparing and gathering his own ideas for a more comprehensive presentation of the whole subject. It provides no detailed apologetic for the orthodox interpretation of *QM* but rather a spring-board from which such an apologetic might usefully be launched. Of topics not referred to in the book, mention might be made of quantum logic. Clearly Scheibe's silence on this subject reflects a nice distinction between the logic of physics and what Finkelstein has aptly referred to as the physics of logic. The translation of the highly technical text is excellent (the only inelegancy noted was in the middle of p. 123) and the number of misprinted symbols and formulae is commendably few. The bibliography is fairly complete although the existence of the excellent English translation of Duhem's *The Aim and Structure of Physical Theory* is surprisingly not referred to. Finally the logical symbolism is somewhat obscure and some symbols are not defined or in one case defined many pages after it was first used. These points should not detract, however, from the final impression that Scheibe has displayed exemplary technical skill in deploying his very detailed arguments, and within its self-imposed limitations this book is undoubtedly an important contribution to the extensive literature on the foundations of *QM*.

M. L. G. REDHEAD
Chelsea College,
University of London

REFERENCES

Araki, H. and Yanase, M. M. [1960]: 'Measurement of Quantum Mechanical Operators', *Physical Review*, **120**, pp. 622–6.

D'Espagnat, B. [1966]: 'Two Remarks on the Theory of Measurement', *Il Nuovo Cimento*, Supplement **4**, pp. 828–38.

Earman, J. and Shimony, A. [1968]: 'A Note on Measurement', *Il Nuovo Cimento*, **54B**, pp. 332–4.

Feyerabend, P. K. [1968]: 'On a Recent Critique of Complementarity: Part I', *Philosophy of Science*, **35**, pp. 309–31.

Feyerabend, P. K. [1969]: 'On a Recent Critique of Complementarity: Part II', *Philosophy of Science*, **36**, pp. 82–105.

Fine, A. I. [1968]: 'Logic, Probability and Quantum Theory', *Philosophy of Science*, **35**, pp. 101–11.

HOOKER, C. A. [1972]: 'The Nature of Quantum Mechanical Reality: Einstein versus Bohr', in R. G. Colodny (*ed.*): *Paradoxes and Paradigms: The Philosophical Challenge of the Quantum Domain*, pp. 67–302.

JAUCH, J. M., WIGNER, E. P. and YANASE, M. M. [1967]: 'Some Comments Concerning Measurements in Quantum Mechanics', *Il Nuovo Cimento*, **48B**, pp. 144–51.

KOCHEN, S. and SPECKER, E. P. [1967]: 'The Problem of Hidden Variables in Quantum Mechanics', *Journal of Mathematics and Mechanics*, **17**, pp. 59–87.

PRUGOVEČKI, E. [1967]: 'On a Theory of Measurement of Incompatible Observables in Quantum Mechanics', *Canadian Journal of Physics*, **45**, pp. 2173–219.

STEIN, H. and SHIMONY, A. [1971]: 'Limitations on Measurement', in B. d'Espagnat (*ed.*): *Proceedings of the International School of Physics 'Enrico Fermi': Foundations of Quantum Mechanics*, pp. 56–74.

VARADARAJAN, V. S. [1962]: 'Probability in Physics and a Theorem on Simultaneous Observability', *Communications of Pure and Applied Mathematics*, **15**, pp. 189–217.

WIGNER, E. P. [1952]: 'Die Messung quantenmechanischer Operatoren', *Zeitschrift für Physik*, **133**, pp. 101–8.

PROPENSITIES AND PROBABILITIES

1 In 1953 or thereabouts, Karl Popper brought forth into the world a new interpretation of probability: the propensity interpretation. According to this view, probability represents something objective and empirical, as it does on frequency or limiting frequency interpretations of probability; but it is a property not of a sequence or class of trials, but rather of a set of circumstances. It is a theoretical property of something (of exactly what we shall have to consider in due course), and more than that it is an explicitly dispositional property. In recent years quite a number of philosophers (Mellor, Levi, Hacking, Gillies, Fetzer, among others) have become persuaded that in principle this is the right approach to interpreting probability. Not only that, in recent years the propensity view of probability has been attributed, retroactively, to such writers as Cramér and Braithwaite and Hempel. There is, to be sure, a 'theoretical counterpart of relative frequencies' view that may be attributed both to Cramér and to Braithwaite (among others) as well, I think, as to Levi; whether this attribution is correct, and whether it is appropriate to characterise it as a brand of propensity theory is a question that would involve me in historical and terminological questions I should prefer to avoid.

There is one terminological matter that is worth mentioning, however. Most of the writers whose ideas I propose to discuss introduce their topic with an explicit disclaimer to the effect that they do not pretend to deal with all the uses of 'probability', but only those uses that are made of 'probability' in scientific hypotheses or statistical laws. It is hard to imagine a statistical law or hypothesis —except of the most elementary kind, hardly worthy of the word 'law'—in which the word 'probability' (or 'chance', or what have you) occurs. Statistical laws are like other quantitative laws: as they are usually expressed, they involve mathematical symbols, and no words, and thus not the word 'probability'. Nor do such laws typically involve even the letter '*P*', nor do they involve the set function '*P*' is sometimes supposed to denote, which in turn might be interpreted as a probability, which in turn might be given one or another sort of ontological or epistemic interpretation. Set functions, satisfying the typical probability axioms, are sometimes encountered in proofs concerning distribution functions and their properties, but this is a slender ground for proposing to talk about 'the meaning of "probability" as that term occurs within the context of statistical laws'. Typically, a statistical law has the form: Q (a quantity, or a random quantity or a random 'variable') is distributed in some set R according to F (a distribution function). For example, 'The weight of adult male frogs of species X is roughly normal (w, d^2)'. One may propose to ***interpret*** this statistical law by reference to phrases involving 'probability'—such as: the law is true if and only if for every k, the ***probability*** of an adult male frog of species K having a weight less than k is $\phi(k)$, where ϕ is the normal distribution function with mean w and variance d^2. This is a curious sort of interpretation, since the variable k is bound by a quantifier *outside* the probability context; we are asked to interpret a relatively straightforward statement about a distribution by means of a continuum of statements about probabilities. Perhaps it can be done and perhaps it is even worthwhile; it is a point we shall return to. I wish to emphasise here only the fact that what seems central to statistical laws in science is the notion of a distribution, rather than the notion of probability. Even the behaviour of the perennial penny is most often described without probability, by saying that the distribution of heads on a sequence of n throws is Binomial with parameters n and 0.5.

Popper introduces the propensity interpretation in the following words:

> 'Every experimental arrangement is *liable to produce*, if we repeat the experiment very often, a sequence with frequencies which depend on this particular experimental arrangement. These virtual frequencies may be called probabilities. But since the probabilities turn out to depend upon the experimental arrangement, they may be looked upon as *properties of this arrangement*. *They characterise the disposition, or the propensity*, of the experimental arrangement to give rise to certain characteristic frequencies *when the experiment is often repeated*'.[1]

As Fetzer and Giere[2] have observed, there is a possible ambiguity in this passage: are probabilities to be identified with virtual frequencies? Or with a dispositional property of the experimental arrangement? This ambiguity recurs later in the same volume; in his reply to those who discussed his paper, Popper writes: 'According to the propensity interpretation, probability statements *may* be said to make assertions about *virtual* frequencies; but they may also be said to make assertions about the experimental situation and the 'tendencies' inherent in it . . .'.[3] The same ambiguity even occurs in his 1959 article, *The Propensity Interpretation of Probability*: '. . . we may interpret "b" [in "$p(a,b)$"] as the name of a (potential or virtual) sequence of events, . . .'.[4]

In the later article he is more explicit than in the earlier that propensities are dispositional properties: 'Like all dispositional properties, propensities exhibit . . .'. But he still fails to see a sharp distinction between associating probabilities with virtual or potential frequencies (in addition to actual ones), and associating them with dispositional properties of experimental arrangements. I think it clear that for Popper the arrangement is fundamental, and that the talk of virtual frequencies is a holdover from Popper's frequentist days. Nevertheless there is a serious question, answered differently by various propensity enthusiasts, as to precisely what it is that probability is to be predicated of. There is also the question of how it is to be interpreted or analysed, or, if it is to be taken as theoretical and primitive, how it is to be given empirical content. Most propensity theorists take probability to be explicitly dispositional; but this solves problems only if both (*a*) we understand how to treat dispositional predicates—which we do not—and (*b*) probabilities are dispositions like other dispositions—which they are not. Although propensity theorists reject the subjectivistic interpretation of probability as a possible interpretation appropriate to physical laws, they nevertheless are not totally insensitive to the arguments for coherence that have been common over the last twenty years. It is obvious that there is some connection between my knowledge of the behaviour of the coin, expressed in a propensity statement, and my willingness to offer certain odds, but no more, on a future occurrence of heads. But what is this connection, and how does it work?

---

[1] Popper [1957], p. 67.

[2] Fetzer [1973]., Giere [1973]. Lack of attention to Giere's work is a serious omission, but I did not have his paper at the time of writing.

[3] Popper [1957], pp. 88–9.

[4] Popper [1959*b*], p. 38.

The following sections will take up these questions one at a time, and review the answers given them by various authors. These questions are: (1) What are probabilities (propensities) attributed to: (2) How are they construed as dispositions? (3) What are the semantics of propensity assertions? (4) How are propensity assertions given empirical content? and (5) How are propensities related to subjective and logical probabilities?

2 I have already quoted Popper as attributing propensities to 'experimental arrangements'. In another place he says that propensities '. . . are not properties inherent in the die, or in the penny, . . . they are relational properties of the experimental arrangement—of the conditions we intend to keep constant during repetition'.[1] But suppose the coin or the die is the only thing we intend to keep constant? Does it make sense to talk this way? It is not immediately clear. We may surely consider throwing the die on various surfaces; we may consider having it thrown on various days of the week, during various seasons, at various times of day. We may consider throwing it by hand, or with a dice cup, or by means of a specially designed machine. We may consider performing the experiment ourselves, or having someone else perform it, or having a number of different people, of different ages, of both sexes, of varying ethnic origins, perform it. It is difficult to see what is held 'constant' except for the die, and the throwing of it. But this is too generous: for we can surely arrange for the die to be thrown by a machine—a very carefully constructed machine—in such a way that it always (or almost always) lands with the six up. It begins to seem as if what is held constant is just those circumstances that determine the chance distribution of outcomes—but of course that would not do, on pain of circularity. Perhaps we are trying to lump too many things together. Perhaps what we are dealing with is not a single propensity assertion, but a whole group of them: the propensity of a six on the die d when tossed under circumstances C_0, the propensity of a six on the die d when tossed under circumstances C_1, . . . , and it is just a coincidence that in many of these circumstances the strength of the propensity turns out to be the same. But now how thoroughly do we distinguish between different

---

[1] *Ibid.*, p. 37.

circumstances? Do we distinguish between machines and people? between one person and another? between a person on one occasion and the same person on another occasion? It certainly would not do to lump all those circumstances that yield the same propensity of six together: one such experimental arrangement would consist of throwing a die and noting the outcome, where we keep track of the ordinal number of the throw and on the prime numbered throws throw a biased die. But this is exactly the kind of sequence that Popper wants to avoid. Yet he writes, '. . . although we may interpret "*b*" [in "*p*(*a*,*b*)"] as the name of a (potential or virtual) sequence of events, we must not admit every possible sequence: only sequences which may be described as repetitions of an experiment are admitted and which may be defined by the method of their generations, that is to say, by a generating set of experimental conditions'.[1]

The answers to these questions lie in Popper's general Philosophy of Science: propensity assertions are general scientific assertions, put forth as bold conjectures, subject to refutation and refinement, just as are other scientific assertions. But in the particular case of propensity assertions, these answers seem anomalous. Popper makes much of the objectivity of propensity statements, and decries at great length the 'subjectivism' of both subjective and logical probability. But here we find that propensities are properties of conditions we ***intend*** to keep constant, of sequences which may be ***described*** as repetitions of an experiment. Thus that of which probabilities are predicated depends on our intentions and the powers of our language—in unspecified ways, note—and since the values of probabilities and the strengths of propensities depend on what it is they are predicated of, these values and strengths in turn depend on our intentions and linguistic abilities. Similar remarks apply to Gillies, who shares Popper's point of view in many respects. Probabilities are objective, and '. . . an underlying deterministic mechanism would refute a probabilistic theory by contradicting the randomness which . . . is demanded by such a theory'.[2] But the randomness referred to is epistemological in character: we determine whether subsequences cause trouble by stipulating the subsequences we will test (designing our tests) before examining the data. 'This seems to be an excellent stipulation which solves the present problem [of criteria for randomness].'[3] Again, whether two events are repetitions depends on 'how we are proposing to describe them'.[4] And swinging back the other way on the next page: *S* is repeatable provided that an 'indefinitely long sequence of repetitions relative to $\mathfrak{S}_s$ [the spacing rule] is in principle (*i.e.* as far as the general laws of nature are concerned) possible'.[5] The real, true, laws of nature—not what we *take* to be the laws of nature.

In both of these approaches there is an unresolved tension between the desire to have probability a wholly objective and theoretical aspect of the world, and the desire to be able to *use* probability in ways that depend on the ways in which sequences are described, our intentions regarding what we shall hold constant, our pre-designation of the tests for randomness that we shall apply, what we know or believe, and so on. This tension need not lead to inconsistency: it is no doubt possible to chart a course between the suggested navigational hazards; the question is merely whether this course takes anyone where they want to go.

There are a number of ways of dealing with the problems posed by these approaches. The most straightforward way is adopted by Hacking: anything whatever can correspond to an 'experimental arrangement', or, as he puts it, a chance set-up. 'A *chance set-up* is a device or part of the world on which might be conducted one or more *trials*, experiments, or observations; each trial must have a unique *result* which is a member of a class of possible results'[1]. Thus if we think of the sequence of possible or virtual results of a chance set-up, any possible sequence is allowed. In particular, the sequence of tosses of a fair die, with a couple of tosses of a biased die thrown in, is a perfectly good sequence for Hacking, in direct and explicit opposition to both Popper and Gillies. This provides a very simple solution to the question of what it is that probabilities or propensities are attributed to: anything satisfying the general description of a chance set up. A toss of a coin (any coin, anywhere, anytime) represents a chance set up; an epidemic at a certain date in a certain city is also a chance set up, though in fact that set of circumstances will never be repeated. The probabilities that are attributed to a chance set up depend on the kinds of trials being performed on it: the chance set up has a propensity to yield a certain result upon being subjected to a certain sort of trial.

---

[1] *Ibid.*, p. 38.

[2] Gillies [1973], p. 135.

[3] *Ibid.*, p. 123.

[4] *Ibid.*, p. 95.

[5] *Ibid.*, p. 96.

This simple solution is bought at a certain cost, of course. While Popper can say that 'the probability of obtaining a six in the third throw made after nine o'clock this morning with this die' is a sixth because this throw may be regarded as a member of a sequence and 'in its capacity as a member of this sequence, it shares in the probabilities of that sequence'[2], the application problem for Hacking is very much more complex. We shall return to the application problem, for it is a problem for all propensity theorists. (As it is for frequency theorists.)

D. H. Mellor provides a solution to the questions we have raised about Popper's propensity approach that in a certain sense is diametrically opposed to Hacking's. That is, while for Hacking almost anything can count as a chance set-up, for Mellor only certain very special circumstances can comprise a chance set-up. A genuine, metaphysical, indeterminism must enter into a chance set-up. If the world is deterministic, there will be no propensities.[3] He shares this view with Gillies, who writes 'If our general approach is right, and if we can correctly assert that at least one probability system exists in reality, then it follows that at least some objective randomness exists in the world, and so complete determinism is shown to be false'[4]. Where Hacking, observing that there is a sense in which a toss of a penny that landed heads may be said to be a trial of the *kind* in which the circumstances always lead to the result heads, says 'far be it from a theory of chance to preclude that august philosophical ogre [of determinism]'[5], Mellor heroically charges forth to grab the ogre by the horns. 'If propensities are ever displayed, determinism is false'[6]. Thus the attribution of a propensity of any strength p to anything is false if determinism is false, and in fact also false if determinism is not false with respect to the instance at hand. It is a little unclear just what determinism is, or where indeterminism may enter—presumably it may enter pretty far back in the historical antecedents of the events in question, since Mellor talks of the diseases and deaths of men as among the 'wide class of cases that might involve propensities'[7], and surely this is not to imply that all death certificates listing the cause of death are false. It does imply that there is, in the history of the deceased, a chance event that lead to his death: *e.g.* when he tossed a coin to choose between two jobs at the age of 20, the result heads ultimately caused his death of malaria at the age of 72.

---

[1] Hacking [1965], p. 13. [2] Popper [1959*b*], p. 29. [3] Mellor [1971], p. 16.
[4] Gillies [1973], p. 133. [5] Hacking [1965], p. 24. [6] Mellor [1971], p. 151.
[7] *Ibid.*, p. 31.

Not only is the domain of application of the notion of a propensity far narrower (particularly if determinism is true) for Mellor than for Hacking, but Mellor attributes propensities to 'more permanent entities' than trials. Indeed, for coins and dice, the propensity is ordinarily attributed to the coin or the die: the physical object in question has a certain propensity to land heads when tossed in standard apparatus. Presumably, it has another (different) propensity to land heads when tossed in non-standard apparatus. Mellor also talks about the propensity of certain apparatus to yield heads when tossing a standard coin.[1] But also, on occasion, the propensity may be ascribed to the chance set-up itself.[2] The motivation for wanting to ascribe propensities to relatively permanent entities (physical objects, *kinds* of trials) is that in due course they will be construed as dispositions; dispositions need relatively permanent abodes whence they can be called forth when their display is provoked, and where they can be supposed to exist even when not being displayed.

3 Let us now look more closely at the disposition in question. Two possibilities present themselves: the disposition of the chance set-up to yield a certain outcome may be construed as either a disposition for a sequence of many trials on that chance set up to yield that outcome, or a disposition for the given single trial to yield that outcome. Or it may be construed as a disposition for that chance set up to yield a certain outcome when subjected to a certain kind of trial, or it may be construed as a disposition for that chance set up to yield a certain outcome with a certain frequency when subjected repeatedly to a certain kind of trial. We may assimilate the latter way of talking to the former by absorbing the kind of trial into the description of the chance set up. (This is appropriate for present purposes only; there are important relations between kinds of trial on the same chance set up that would be lost sight of if we adopted this approach generally.) Observe that our decision on this matter can be quite independent of what it is we have decided to attribute the disposition to: whether an experimental arrangement, a chance set-up, or a physical object. Popper never quite made up his mind with regard to the disposition that is propensity. He often talks of potential or virtual sequences, and the relative frequencies therein; but he also says '. . . we now take as fundamental *the probability of the result of a single experiment* . . .'[3] and asserts that '*the propensity distribution attributes weights to all the possible results of the experiment*'[4]. It is easy to overlook this distinction, because in either case we may be confident that in the long run frequencies will approximate chances. But as Fetzer has pointed out[5] there is both an important logical difference and an important metaphysical difference. If we take the disposition to characterise a single trial, it then follows from Bernoulli's theorem, with limiting probability one, that in a long enough sequence of independent trials, the frequency of outcomes of a given kind will very nearly equal the strength of the propensity for each of those trials to produce that outcome. On the other

---

[1] *Ibid.*, p. 63, 75.

[2] *Ibid.*, p. 67.

[3] Popper [1957], p. 68.

[4] *Ibid.*, p. 68.

[5] Fetzer [1973].

hand, if we take the disposition to characterise the (possible, or virtual) sequence of trials of that kind, then it is a matter of logical necessity that the long run frequency will be equal to the strength of the propensity. The metaphysical difference is probably even more important than the logical difference to those who wish to attribute the disposition to the single trial. We shall return to this matter shortly.

Of propensity theorists, Hacking is most unambiguous about interpreting propensities as dispositions to exhibit certain long run behaviour. The propensity '. . . is a dispositional property of the coin: what the long run frequency is or would have been'[1]. Levi seems to adopt the same stance, and in his review of Mellor's book accuses Mellor of failing 'utterly' to show why the treatment of single case chance is essential to an understanding of statistical truth. He says that neither he nor Hacking assumes that probability in the single case is statistical or chance probability. On the other hand, Levi has made it clear (in personal communication) that he does not want to be understood as a long-run propensity theorist: certain kinds of trials on a chance set up (long run trials) are of peculiar evidential importance, but statements concerning them do not provide even a partial interpretation of propensity. Gillies cites Peirce with approval for interpreting probability in analogy with habit, and wholeheartedly endorses Peirce's distinction 'between the probability of the die as a dispositional quantity, a 'would be', on the one hand and an occasion [an infinite sequence of throws of the die] that would bring out the full consequences of the 'would be' on the other'[2]. Thus he, too, appears to be committed to the long-run hypothetical-frequency view of propensities—combined, in his case, with a requirement of independence.

Fetzer and Mellor, on the other hand, want to attribute dispositions to the individual trials. Fetzer is motivated in part by a desire to solve the problems of the single case application of probabilities (provided determinism is false, there is a precise physical objective probability that the next toss of this coin will yield heads; this probability is independent of what anyone knows or believes; it is independent of language and intention, *etc.*). More importantly, perhaps, he is motivated by a desire for explanation which he feels is satisfied by the attribution of propensities to individual trials. We observe that certain statistical patterns occur in the world. '. . . the problem, therefore, is to provide an explanation for the occurrence of these patterns . . . the propensity interpretation provides a theoretical basis for accounting for these patterns in terms of the ***system's initial conditions***, since the occurrence of actual frequencies is explained by reference to the dispositional tendencies that generate them . . .'[3]. Motivation aside, Fetzer construes propensities as unlike other dispositions. While a disposition like fragility invariably displays itself on a test condition, a propensity to yield heads does not display itself in an interesting way on a single trial, but only as a relative frequency on a number of trials. But this is just because the disposition in question *is* statistical rather than universal.

Mellor, on the other hand, wants a disposition to be displayed on *every* occurrence of its test condition. Thus for him the propensity is displayed on every trial—and indeed part of the reason for his wishing to attribute the disposition to the single trial may be his attitude toward dispositions in general: namely that if something has a disposition, and the circumstances which call forth a display of that disposition obtain, then the disposition *must* be displayed. The salt *must* dissolve in the water. A disposition to yield heads half the time cannot be displayed directly on a single toss; there is no way heads can come up half the time on a single trial. Furthermore, we wish to speak of the probability of events in general, such as the event of getting an odd number on a roll of a die, and not merely of the probability of results. But a given trial can have only one result, and we may be asked what possible connection there is between the disposition of a die to yield a one (which it did) and the disposition to yield an odd number, which is what we are interested in.

---

[1] Hacking [1965], p. 10. [2] Gillies [1973], p. 145. [3] Fetzer [1973], p. 10.

In any event, Mellor wants to regard the display of propensity as the same on every test of it, just as the display of salt's solubility occurs on every test of it. What is displayed is not the outcome (which of course must vary from case to case) but the chance distribution. Of course the chance distribution may be got at through testing—we make a large number of trials, and so on—but it is the propensity that is scientifically and metaphysically important. It is pervasive and permanent—while the chance distribution (itself only indirectly observable) varies according to the circumstances under which the propensity manifests itself. 'The display of a propensity is the chance distribution over the *possible* results of *the* appropriate trial.'[1]

It seems clear that both Mellor and Fetzer are quite serious about attributing to chance a fundamental role in their picture of the world. This comes out both in the restriction to non-deterministic kinds of events, and in the ascription to single trials of the dispositional property. Unlike Mellor, Fetzer is quite willing to say that the propensity displays itself on a given trial in the outcome on that trial—though not in the same way on every trial, like the disposition of solubility. Nevertheless, that disposition is in the world, and ineradicably in the world. It is not a reflection of our epistemological state (obviously), nor of the scientific language we happen to talk, nor of how we describe things. Certain things take place by chance; propensities exist, whether or not they are ever tested or called forth or displayed. Whether one wishes to predicate the propensity of a kind of trial, and use that as an explanation of the distribution of chances that is characteristic of a single trial, or whether one wishes to predicate propensities of the single trial directly, there is a special characteristic of the universe that we are capturing if we use propensity and probability talk correctly. When we use this kind of talk correctly, we simply *are* attributing a certain characteristic to certain events.

One of the attractions of propensity theory is that it allows of—indeed, seems to have a propensity to call forth—wild metaphysical speculation in a context in which one is freed of any constraint to provide careful metaphysical arguments. Thus Fetzer claims that there is a *real connection* between the two terms of a statistical law, of which the causal relation is merely a special case. Gillies, even though he does not accept a single case propensity interpretation of probability, feels that if any probability assertions are true, determinism is false. Mellor makes similar assertions in several places, and feels himself forced, in order to take account of the use of statistics in the insurance business, to draw a strained analogy between physics and medicine. The chances of decay of an atom, he says, 'are derivable from the theory as they [the chances of death] are from the theory of physiological aging'[2]. But there *is* no theory of physiological aging in

[1] Mellor [1971], p. 70. [2] *Ibid.*, p. 98.

the sense that there is a theory of radioactive decay. If we could identify people chemically and physically so as to put them in a small number of classes, each with a different half-life, and that is *all* we could do, then we would be home free. But precisely in distinction to the decay of radioactive elements, we can often *identify* the cause of a person's demise; and furthermore, perhaps naively, we suppose in conformity to the supposition of the classical English Coroner of detective stories, that whenever there is a death there is an identifiable specific cause for that death. The theory of aging, if there were to be one, therefore would seem precisely to exclude treatment on the model of the theory of radioactive decay. In certain cases Mellor seems to feel the same way; he remarked earlier 'A man may obviously be so [*sic*] assassinated that his death is not a matter of chance at all'[1].

4 One way in which to get a little clearer as to what is entailed by these various interpretations of statistical propensities is to take a look at the semantics that might be provided for them. At the same time, this will help us to see somewhat more clearly the connection between classical frequentist views of probability and propensity views. This is of interest, since one thing that all propensity theorists seem to have in common is a propensity to attack such arch-frequentists as von Mises.

Let us consider a first order predicate calculus with identity and operation symbols, and with enough mathematical machinery to allow the expression of statistical hypotheses. A model M for this language consists of an ordered couple $\langle \mathscr{U}, \mathscr{A}_{\mathscr{U}} \rangle$, in which $\mathscr{U}$ is a set—the set of entities in the universe—and $\mathscr{A}_{\mathscr{U}}$ is a function whose domain is the set of non-logical predicates and operators of the language L. To a one-place predicate $\mathscr{A}_{\mathscr{U}}$ assigns a subset of $\mathscr{U}$; to a k-term relation a subset of $\mathscr{U}^k$; to a 0-place operation symbol (name) a member of $\mathscr{U}$ (possibly the empty set, which we suppose to be a member of every $\mathscr{U}$, in case nobody else in $\mathscr{U}$ answers to that name); to a k-place operator, a function from $\mathscr{U}^k$ to $\mathscr{U}$ (possibly having as value the empty set, in case $\mathscr{U}$ is not closed under the operation in question). We suppose that the functions $\mathscr{A}_{\mathscr{U}}$ are compatible in the following sense:

(*i*) for every predicate π in the language,

$$\mathscr{U}_1 \subset \mathscr{U}_2 \rightarrow \mathscr{A}_{\mathscr{U}_1}(\pi) \subset \mathscr{A}_{\mathscr{U}_2}(\pi)$$

(*ii*) for every k-place operation expression ω in the language,

$$\mathscr{U}_1 \subset \mathscr{U}_2 \rightarrow \bigwedge_{x \in \mathscr{U}^k} (\mathscr{A}_{\mathscr{U}_1}(\omega)(x) = \mathscr{A}_{\mathscr{U}_2}(\omega)(x) \vee \mathscr{A}_{\mathscr{U}_1}(\omega)(x) = 0)$$

We suppose that the axioms (definitional and otherwise) of the language are true in every model. The actual world we take to be a given model M^*, with the actual universe of entities $\mathscr{U}^*$, and the actual assignment $\mathscr{A}_{\mathscr{U}^*}$. We shall have also to deal with a few other models. The set of lawful models, $\mathscr{L}\mathscr{M}$, is to be the set of models in which the set N of non-statistical statements representing the set of physical laws are true. The set of future models, $\mathscr{F}\mathscr{M}$, is to be the set of models in which the set O of statements representing the set of our observations up to the present are true.

---

[1] *Ibid.*, p. 90.

On the basis of $\mathscr{A}_{\mathscr{U}}$, we build up an extension assignment, ϵ_M, for the model $M = \langle \mathscr{U}, \mathscr{A}_{\mathscr{U}} \rangle$. A model M_1 is an *extension* of a model M_2 just in case the universe of M_2 is included in the universe M_1, and the extension assignments of M_1, restricted to the universe of M_2 in an obvious sense, are the extension assignments of M_2. M_1 is an *x-maximal* extension of M_2 if it is an extension of M_2 and x is a term of L and either what ϵM_1 assigns to the term x is already infinite or there is no extension of M_2 that assigns to x a set that properly includes in $\epsilon M_1 (x)$.

We need one more bit of notation: we suppose that there is a three place predicate S in the object language where $S(x,y,z)$ has the intuitive meaning: among the objects in the reference sequence x, the frequency of members of y is z. More precisely, $S(x,y,z)$ is to hold just in case:

(*i*) x is a non-decreasing, infinite, sequence of finite sets, *i.e.*
x is a function with domain ω such that x_i is finite, and if
$i<j$, $x_i \subset x_j$.

(*ii*) y is a set

(*iii*) z is a real number

(*iv*) $$\lim_{i\to\infty} \frac{\text{the number of members of } y \cap x_i}{\text{the number of members of } x_i} = z$$

We are now in a position to characterise semantically those objective theories of probability to which we have referred—and some others as well.

Only Bertrand Russell[1] and L. Sklar[2] seems to qualify as extensional finite frequentists. For them a statistical statement $S(x,y,z)$ is true just in case it is true in the actual world (M^*), and the number of members of $\cup \mathscr{R}x$—the union of the range of x—is finite. If there are ten balls in an urn and three are black, and we take x to be the infinite sequence each of whose terms is the set of ten balls in the urn, and y to be the set of black things, then $S(x,y,0.3)$ says exactly that the probability of a ball in the urn being black is 0.3, and this is true in the world we have described.

Propensity theorists have correctly pointed out a number of serious difficulties with this view. For example, if a newly minted coin is to be tossed once and then destroyed, how can the probability of heads on this extensional finite frequency view be other than 0 or 1? Levi's recent reply to Sklar[3] details a number of other objections.

Reichenbach may be construed as an extensional finite-or-infinite frequentist —he abandons the requirement that $\cup \mathscr{R}x$ be finite. But this is a relatively unimportant difference—we can always restrict our concerns to the first 10^{100} trials, or to those individuals encountered in a region 10^{100} light years in radius and during the next 10^{100} years.

We have already run out of extensional frequentists. Even von Mises—one of the prime targets of any propensity theorist—makes it perfectly clear that his probability statements are theoretical and ideal. Of course all actual sequences are finite in *this* world; but we are talking (according to von Mises) of *ideal* or *hypothetical* worlds in which sequences of trials are unending. This leads us to consider *possible* worlds, and an intensional predicate S^*. Informally, $S^*(x,y,z)$

---

[1] Russell [1948].
[2] Sklar [1940].
[3] Levi [1973].

means that the propensity or disposition of the sequence x to produce members of y is z. More formally, to obtain truth conditions for $S^*(x,y,z)$ for von Mises, we must consider possible worlds in which (in our language L) there is no way to characterise a winning gambling system in a finite manner. Thus for von Mises,

> $S^*(x,y,z)$ is true in the actual world just in case every x-maximal extension of a member of $\mathscr{F}\mathscr{M}$ (the set of future worlds)—or alternatively of $\mathscr{F}\mathscr{M} \cap \mathscr{L}\mathscr{M}$ (the set of lawful future worlds)—is such that
>
> (*i*) $S(x,y,z)$ is true in it
>
> (*ii*) The assignment of $\cup \mathscr{R}x$ in it is infinite
>
> (*iii*) There is no finite way to characterise a winning gambling system in that world.

In effect, then, to make an intensional, lawlike, probability statement, for von Mises, is to say something about every possible world which is large enough, and which satisfies two further constraints.

Most statisticians, like Cramér and Kolmogorov, don't regard the further constraints as important; we may for all practical purposes take the 'standard' hypothetical frequency interpretation of S^* to be:

(A) $S^*(x,y,z)$ is true in the actual world just in case
$S(x,y,z)$ is true (nearly true) in every (almost every)
x-maximal extension of a member of $\mathscr{F}\mathscr{M}$.

Let us now consider propensity theories. According to these theories, we interpret propensities as dispositions. The ordinary possible-world analysis of a disposition predicate is that it applies to something just in case in every possible world in which that thing is subject to a specified test condition, it responds in the appropriate way. 'This lump of sugar is soluble' is true in this world, for example, because in every *possible* world in which it is put in water (enough water) and is shaken up (for long enough) it dissolves. To assert that something is soluble is to constrain the set of possible worlds; it is to rule out some worlds as impossible.

The dispositional analysis of probability offered by Peirce and Hacking seems to admit of a perfectly analogous treatment. To assert $S^*(x,y,z)$ is just to constrain the set of possible worlds in such a way that in every maximal extension of a world in $\mathscr{F}\mathscr{M}$, the extensional statement $S(x,y,z)$ is true. The semantic treatment of propensity statements seems to be indistinguishable from the semantic treatment of the hypothetical frequency interpretation, (A).

The same truth conditions for statements of the form $S^*(x,y,z)$ are necessary for Gillies and for Popper. However, they are not sufficient. Popper would impose further constraints on the term x—to rule out sequences of dice-throws, for example, in which every prime numbered toss is performed with a biased die. Gillies would impose two further requirements. For one thing, he wishes to consider only lawful possible worlds—members of $\mathscr{F}\mathscr{M} \cap \mathscr{L}\mathscr{M}$. 'A set . . . of conditions is repeatable provided that an indefinitely long sequence of repetitions . . . is in principle (*i.e.* as far as the general laws of nature are concerned) possible.'[1] His condition regarding randomness embodied in his requirement of

---

[1] Gillies [1973], p. 96.

independent repetitions is not altogether clear, but from his comments about determinism it is clear that a *consequence* of this requirement would be that there be more than one maximal extension of any possible world in $\mathscr{F}\mathscr{M} \cap \mathscr{L}\mathscr{M}$. It *might* be that the requirement is this: given that $S^*(x,y,z)$ and that $\langle \mathscr{U}, \mathscr{A}_{\mathscr{U}} \rangle$ is a possible world in $\mathscr{F}\mathscr{M} \cap \mathscr{L}\mathscr{M}$, there must be a set of n extensions of $\langle \mathscr{U}, \mathscr{A}_{\mathscr{U}} \rangle$, each containing one additional object in the extension of $\cup \mathscr{R}x$, such that in zn of them the additional object also belongs to the extension assigned to y. (This would be awkward if z were irrational, but no matter.) Note that in principle this approach appears to yield a semantics almost indistinguishable (except for the restriction to *lawful* possible worlds) from that suggested for von Mises' probabilities.

Finally, we come to Mellor and Fetzer, who have something genuinely incompatible with classical frequency views to offer. Fetzer wishes to attribute probabilities specifically to the individual members of $\cup \mathscr{R}x$. For example, if a is a chance set up belonging to a set in the reference sequence x, we are required to attribute to a itself a probability z of yielding an outcome belonging to y. Thus as a first step we must have in the language L a predicate $Pr_{z,y}$ which holds of all and only those objects that have a propensity z to yield a member of y. To assert $S^*(x,y,z)$ is thus to assert a good old nomological conditional:

(B) In every possible world, every member of $\cup \mathscr{R}x$ has the property $Pr_{z,y}$.
But what is this property? Fetzer takes $Pr_{z,y}$ to be a statistical dispositional property; the nomological conditional is analogous to:

> In every possible world, every lump of sugar is soluble.

A statistical dispositional property is one that does not exhibit the same result under every test conditions, but only with a certain frequency. It is a part of the nature of statistical disposition properties that they obey the law of large numbers. That is, given a lot of objects with the property $Pr_{z,y}$, that lot will, with probability 1, come to exhibit a frequency of members of y that is close to z, as the size of the lot increases. In terms of our models:

> $S^*(x,y,z)$ is true in the actual world, just in case (B) is true in the actual world; and the latter is true in the actual world just in case, in almost all x-maximal extensions of a world belonging to $\mathscr{F}\mathscr{M} \cap \mathscr{L}\mathscr{M}$, $S(x,y,z)$ is true.

Shades of Cramér, except for the introduction of the predicate $Pr_{z,y}$! Now it may be that in some metaphysical system these predicates will serve an explanatory function, but it is altogether unclear what explanatory function they serve in the context of ordinary statistics that is not served by the statistical statement $S^*(x,y,z)$ as it is interpreted in the standard hypothetical frequency way. Fetzer in addition requires ontological randomness—observe the restriction to lawful possible worlds. Perhaps, as in the case of Gillies, this can be explicated as the requirement that there exist an appropriate number of possible worlds with appropriate characteristics.

Mellor's semantics (if I am right) is even more complicated. Mellor takes the dispositional predicate to apply to some permanent entity existing in each of the trials comprising $\cup \mathscr{R}x$. Thus if x_i is the set consisting of the first i tosses of coin j (in a standard apparatus), the propensity to produce heads half the time is attributed to j. This is a permanent property that j has. In the language L we must therefore have a predicate $Pr_{x,y,c}$ which holds of all and only those objects that, subjected to test conditions c, exhibit a chance distribution for y characterised by z. This chance distribution is exhibited on *every* test of j under conditions c, that is, on every member of $\cup \mathscr{R}x$. There appears to be no difference between the display of this chance distribution by a trial, and the possession by the trial of Fetzer's dispositional property $Pr_{z,y}$. Thus in the framework of our models, we have:

> j has the property $Pr_{z,y,c}$ just in case, in every possible world, j, subjected to a test under conditions c, will exhibit $Pr_{z,y}$

Now x_i may be taken as the first i trials of j under conditions c. It follows, then, as for Fetzer, that every member of $\cup\mathscr{R}x$, in every possible world, will exhibit (display) $Pr_{z,y}$. And although Mellor does not regard $Pr_{z,y}$—the property that a trial has when it displays a certain chance distribution—as a dispositional property, it is a property that can be measured, thus providing an indirect measure of the propensity $Pr_{z,y,c}$ attributed to j. Mellor regards the problems surrounding this form of measurement as involving such obscure and questionable notions as confirmation theory, but in no more an invidious way than in any other form of measurement. It is none-the-less clear that—on any reasonable theory of measurement—it will turn out that

> In almost all x-maximal extensions of a world in $\mathscr{F}\mathscr{M} \cap \mathscr{L}\mathscr{M}$, $S(x,y,z)$ will be true.

Just as Fetzer's dispositional predicate $Pr_{z,y}$ failed to add anything to $S^*(x,y,z)$ —since it applied to *every* trial belonging to $\cup \mathscr{R}x$—so it is even harder to see how Mellor's interpolation of yet a further property into the machinery serves a useful function. Mellor, like Fetzer, suggests that it is a question of explanatory coherence: 'Propensities admitted in scientific theory must thereby be connected to other propensities in whose terms they admit of explanation.'[1] But this suggestion is not much clarified, even by the following down-to-earth chapter on 'Half lives and the force of mortality'.

Partly as a matter of self-indulgence, and partly because I suspect that it may interest some readers, I shall close this section with a brief summary of my own view of statistical laws. I do not regard it as a propensity interpretation of probability, because neither the term 'probability', nor the letter 'P', nor anything reminiscent of either of them, occurs either in explicans or in explicandum. (Recall the diatribe regarding diction of pages 1 and 2.)

I take as fundamental the locution: in the reference sequence x, the random quantity y is distributed according to one of the distribution functions of the set z. In symbols, I write $D(x,y,z)$. Formally, $D(x,y,z)$ holds just in case,

(*i*) x is a non-decreasing ω-sequence of finite sets (as before)

(*ii*) y is a function from $\cup \mathscr{R}x$ to real numbers (or, more generally, n-sequences of real numbers)

(*iii*) z is a set of statistical distribution functions (a distribution function being a purely mathematical function with certain formal properties)

---

[1] Mellor [1971], p. 71.

(*iv*) There exists a z^* in z such that for every real number (n-sequence of reals) s,

$$\lim_{i\to\infty} \frac{\text{number of members of } x_i \text{ such that their } y\text{-values are less than } s}{\text{number of members of } x_i} = z^*(s)$$

Note that the earlier statements of the form $S(x,y,z)$ can easily be defined in terms of $D(x,y,z)$, but not conversely: a statement of the form $D(x,y,z)$ can only be represented (sometimes) by an infinite number of statements of the form $S(x,y,z)$. Observe furthermore that in real science one tends to be interested in distributions rather than in simple frequencies.

Statements of the form $D(x,y,z)$ are, like statements of the form $S(x,y,z)$ purely extensional statements whose truth conditions in any possible world are perfectly straight-forward. We often wish to treat statistical laws intensionally. We do so by introducing the intensional predicate D^*, and providing it with a possible world semantics as before.

$D^*(x,y,z)$ is true in the actual world just in case
$D(x,y,z)$ is true in every x-maximal extension of a member of $\mathcal{F}\mathcal{M}$

We may restrict our concerns to lawful worlds, or we may restrict our concerns to 'accessible' worlds, or we may restrict our concerns to any combination of these. In any event, the principle seems reasonable: it is distribution statements that are fundamental, and they are true in an intensional sense just in case they are true in an extensional sense in worlds that are large enough.

5 In *Scientific Explanation* Braithwaite offered an explanation of the meaning of statistical laws in terms of a rule for rejecting them in the face of contrary evidence. The rule was not construed as definitive: a given rejection could, in principle, always be overthrown by new evidence. Nevertheless, Braithwaite contended that the 'possibility of an unending sequence of rejections'[1] sufficed to give empirical content to the statistical law. Hacking, Levi, and Gillies follow this same general ploy: attributions of propensities are given empirical content in terms of rules for acceptance and rejection. Just as the dispositional property of being soluble is given meaning to the correlative rule: put it in water and see what happens—which is, as philosophers unrealistically regard these things, an infallible truth-preserving rule—so the dispositional property of yielding heads half the time is given meaning by a rule to the effect that if on 10,000 tosses too many or too few yield heads, the dispositional statement is to be rejected.

What is at issue is tied up with and confused by a whole gaggle of philosophical issues that are part of the legacy of the past fifty years of meaning analysis. Some people regard it as a clearcut consequence of all this analysis that dispositions *must* display themselves under appropriate test conditions. Some people regard it as a clearcut requirement of empirical meaningfulness that there *must* be at least necessary conditions for the application of a meaningful term. Some people regard it as a clearcut requirement of rules of inference that *must* they be truth-preserving. All of these clearly understood necessities are thrown into doubt by the attempt to take statistical knowledge and statistical inference seriously. It is clear that for Mellor, for Popper, for Fetzer, as well as for Levi and Hacking and

---

[1] Braithwaite [1953], p. 162.

Gillies, some form of non-truth-preserving inference must be taken to provide the empirical content of probability assertions. Only the last mentioned three philosophers, however, descend into the pit to grapple with the basic problem—inevitable however we regard dispositions and their displays, inevitable whatever we attribute propensities to—of coming to grips with the details of statistical inference.

Popper has noted that for statistical statements the principle of falsifiability must be construed in a novel way.[1] Mellor dismissed the problem as in principle not different from that of measuring any other inexact dispositional property. His discussion of the matter[2] is somewhat confused, and often incorrect in its reporting of statistical practise. For example, in the context of a discussion of Neyman-Pearson Hypothesis testing, he writes 'No doubt $(P_a(H) = 0.5)$ cannot be tested directly'.[3] But this is exactly the sort of hypothesis the theory of testing statistical hypotheses was designed to account for tests of. He suggests that tests of hypotheses of the form $k < P_a(H) < j$ *can* be tested, but that any 'test criterion will leave a finite probability of rejecting a true statement of [that] form . . . '.[4] Of course on the classical statistical view there is no such probability: there is only the conditional probability of rejecting H by the criterion, *given* that H is true. Trivial details, to be sure, but they suffice to show that he does not take seriously the basic problems of statistical inference, even though one would think that they had to be dealt with *prior* to supposing, as Mellor does, that we understand about measurement.

Hacking, Gillies, and Levi do take these problems seriously, and their proposals are worthy of far more serious consideration than can be given them here. Hacking offers a principle, *The Law of Likelihood*, in terms of which he develops a theory of statistical support. That a statistical hypothesis can receive support provides it with empirical content. There is one sort of question that receives no answer: given a toss of a coin, of what *kind* of chance set-up shall we regard it as a trial? It is also characteristic that we may only test statistical hypotheses against explicit alternatives. Where do we draw these alternatives from? Background knowledge, intuition, common sense. Much the same is true for Levi, who provides a principle of statistical inference in the form of his Rule A.[5] The rule applies, however, only when we have a list of relevant alternative hypotheses; this list, so far as his work to date is concerned, must be obtained from pre-systematic considerations. Gillies formulates an *FRPS*—a Fundamental Rule of Rejection for Probability Statements—which involves explicitly the magic number of psychology and social science, 0.05. He, too, is ultimately forced to refer to common sense, and 'the principle that we should prefer tests based on statistics which either measure quantities of practical interest or are closely related to such quantities'.[6] Now despite the handwriting around the periphery, all three of these views have important practical consequences for the conduct of statistical inference that should be considered seriously. In the present place those details are not very relevant; what is relevant is that on each view the empirical meaningfulness of statistical statements is tied to a non-truth-preserving rule of acceptance (or rejection); and that on each view the rule involves certain practical or factual assumptions or presuppositions.

---

[1] Popper [1959*a*], p. 146.
[2] Mellor [1971], ch. 6.
[3] *Ibid.*, p. 102.
[4] *Ibid.*, p. 102.
[5] Levi [1967], p. 86.
[6] Gillies [1973], p. 225.

6 There is one more problem that an interpretation of probability should deal with: the problem of application. For Fetzer the problem is straightforward, though his solution is of little practical help. It is individual trials that are the bearers of probabilities, and the correct probability to attribute to a particular trial is the one that, in the metaphysical scheme of things, is true of it. If the world is deterministic, then, of course, all attributions of probability are equally wrong.

For Hacking and Levi, the problem is also fairly straightforward, in principle. The Law of Likelihood, and Rule *A*, will serve as well (perhaps better?) to guide the attribution of probability to a particular instance on the basis of known probability laws, as to relate our choice of probability laws to the observation of frequencies. Neither rule, however, has much to offer in cases in which we have conflicting statistical laws: *e.g.* when we know that a toss is performed in an apparatus biased toward tails with a coin that is biased toward heads. That is, neither comes to grips directly with the problem of the choice of a reference class.

Gillies reveals here the tension between the theoretical and the practical alluded to earlier. At one point, when he is talking of the probability of getting an ace of spades in an experiment which consists of drawing the cards one at a time from a deck, without replacement, he says that '. . . if we don't know the previous results then we ought to say we just don't know the value of the singular probability . . .'.[1] Here the 'correct' probability is that metaphysical one which depends on the number of cards left in the deck and on whether or not the ace of spades is left in the deck, whether we know these things or not. Later on, however, he regards the choice of a reference class as essentially a pragmatic problem: 'The reasons for preferring one class to another are given by considerations external to the frequency theory'. And, as he admits, external to his own theory, too.[2]

Of those dealt with in this particle, Mellor takes most seriously the problem of the application of probabilities. He writes, 'Single case application is so central to the use of probability statements . . . that no account of their meaning can be acceptable which rules it out'.[3] He regards this as a problem for both frequency and logical relation theorists: 'Their attempts founder on the problem of setting a non-arbitrary limit to the amount of evidence, or closeness of specification of the reference class . . .'.[4] With regard to logical theorists, the claim is false. After all these years it is hard to believe that anyone could so misconstrue the principle of total evidence. That principle is not without its problems—how does evidence come to *be* evidence?—but there is no obvious problem about taking everything I know at a given time and regarding that as the total evidence, nor is there any problem about taking everything I know at a given time, and using that as a basis for evaluating the *expected* cost or profit of getting more evidence. But Mellor is quite right in pointing out the inadequacy of the casual (merely 'methodological') rules suggested by Reichenbach and Salmon—and also by Gillies when he talks of the problem of the choice of a reference class. But after all this Mellor goes on to develop a theory according to which *all* probability statements may be false; according to which we have no way of *determining* the correct probability to apply to the single case—if there is one; in which not even the vague and inadequate methodological rules for choosing a reference class supplied by Reichenbach and Salmon have counterparts.

---

[1] *Ibid.*, p. 152.

[2] *Ibid.*, p. 156.

[3] Mellor [1971], p. 54.

[4] *Ibid.*, p. 54.

This is the self-imposed dilemma of all propensity theories: they are (in part) inspired to reject frequency theories for the sake of a practical problem—the single case—and then, in the attempt to account for the probability of the single case, they find themselves in a position where the single case can be accounted for—but only in such a way that the concerns that motivated a departure from a frequency view are as frustrated as ever. Mellor's suggestion, like Fetzer's, that there is a real and true probability which is really and truly applicable to the particular case, is a council of perfection. I want to know how to apply the knowledge I happen to have to that particular case. The deeply metaphysical propensitist has less advice to offer me than many a frequentist.

7 I conclude with what I hope is a handy table of the positions I take several leading propensitists to hold on the issues we have discussed.

Disposition attributed to:

chance set-up	Hacking, Levi
uniform trials of experiment	Popper
random sequence	Gillies
an object: die, or tossing apparatus, (or concretely described kind of trial)	Mellor, Peirce
particular trials	Fetzer

Nature of disposition:

long run relative frequency	Peirce, Hacking, Levi
long run + randomness or independence	Gillies, Popper (?)
kind of trial, or physical object; propensity always displayed in chance distribution (characterised on *each* trial)	Mellor
trial on chance set-up—propensity only displayed in frequency distribution over long run	Fetzer

Semantics:

disposition directly revealed in any large enough possible world	Hacking, Levi, Gillies, Popper, Peirce
disposition revealed indirectly with high probability in a possible world	Fetzer
disposition displayed directly in every possible world; display, in turn, measurable in any large enough world	Mellor

Empirical meaningfulness

Rule A	Levi
Law of likelihood	Hacking
FRPS	Gillies

Application problem

Law of likelihood	Hacking
Rule A	Levi
true probability or extrasystematic	Gillies
true probability	Mellor, Fetzer

HENRY E. KYBERG
University of Rochester

REFERENCES

Braithwaite, R. B. [1953]: *Scientific Explanation.*
Fetzer, J. H. [1971]: 'Dispositional Probabilities', in *Boston Studies in The Philosophy of Science*, **8**, pp. 473–82.
Fetzer, J. H. [1973]: 'Physical Probabilities: a Conceptual Dilemma', *unpublished.*
Gillies, D. A. [1973]: *An Objective Theory of Probability.*
Giere, R. [1973] 'Objective Single Case Probabilities and the Foundations of Statistics', in P. Suppes *et. al (eds.): Logic, Methodology, and Philosophy of Science* **4**, North-Holland, Amsterdam.
Hacking, J. [1964]: 'On the Foundations of Statistics', *British Journal for the Philosophy of Science*, **15**, pp. 1–26.
Hacking, I. [1965]: *The Logic of Statistical Inference.*
Hacking, I. [1967]: 'Slightly more Realistic Personal Probability', *Philosophy of Science*, **34**, pp. 311–25.
Levi, I. [1967]: *Gambling with Truth.*
Levi, I. [1969]: 'Are Statistical Hypotheses Covering Laws?' *Synthese*, **20**, pp. 297–307.
Levi, I. [1973]: '. . . But Fair to Chance', *Journal of Philosophy*, **70**, pp. 52–55.
Mellor, D. H. [1971]: *The Matter of Chance.*
Popper, K. R. [1957]: 'The Propensity Interpretation of the Calculus of Probability, and the Quantum Theory', in S. Korner (*ed.*): *Observation and Interpretation.*
Popper, K. R. [1959]: *The Logic of Scientific Discovery.*
Popper, K. R. [1956]: 'The Propensity Interpretation of Probability', *British Journal for the Philosophy of Science*, **10**, pp. 25–42.
Russell, B. [1948]: *Human Knowledge.*
Sklar, L. [1940]: 'Is Probability a Dispositional Property?' *Journal of Philosophy*, **67**, pp. 355–66.
Sklar, L. [1973]: 'Unfair to Frequencies', *Journal of Philosophy*, **70**, pp. 41–52.
Watling, J. [1969]: 'Chance', *Aristotelian Society, Supplementary volume*, **43**, pp. 34–48.

MAURICE CLAVELIN ON GALILEO'S NATURAL PHILOSOPHY*[1]

Maurice Clavelin has given us, in the present monograph, a valuable work: by far the best account known to the reviewer of the *oeuvre* of Galileo; and an exemplar of a type of study in the history of thought that is, in the reviewer's opinion, both highly desirable and all too rare. The aim and scope of his treatise are indicated by the author in his foreword as follows:

> This book has as its object the Galilean *oeuvre* in its historic singularity. It is constructed about a specific question: how did Galileo create the

---

* Work on this article was supported in part by the National Science Foundation.

[1] Review of Maurice Clavelin, [1968]: *La Philosophie Naturelle de Galilée. Essai sur les Origines et la Formation de la Mécanique Classique*, Paris: Librairie Armand Colin.

> modern science of motion? Starting from this question, we wish to determine, as exactly as possible, in respect of what problems, by means of what concepts and what methods, classical mechanics was able to take shape.
>
> We have not attempted any systematic reconstitution of the intellectual and social context in which the Galilean science developed. We do not claim, of course, that this reconstitution would be devoid of interest. . . . But that path, in our view, cannot lead to what is essential. The science of Galileo is first and foremost the passage from one conceptual universe to another conceptual universe, the replacement of one explanatory ideal by another explanatory ideal, a novel alliance of reason and the real; to comprehend in its right dimensions this prodigious overthrow, to evaluate its bearing and signification, there is in fact only one possible attitude: to place oneself deliberately inside the work, to attempt to reconstruct it in its organization and its complexity—without omitting, should such occur, its difficulties, its lacunae, even its contradictions. . . .
>
> The same intent—to appraise in all its originality the contribution of Galileo—has motivated the attention that we accord to the traditional analysis of motion. Its systematic reconstruction seemed necessary for two reasons. If it is true that Galileo's science marks the advent of a new conceptual universe, one will certainly not be able to prove this better than by setting up, at the outset of the exposition, that universe whose definitive collapse the new science provoked. On the other hand, the farther we advanced in our work the clearer it became, not only that the traditional physics which Galileo echoes cannot be reduced to the letter of Aristotle, but also that certain of its themes, introduced in the fourteenth century, . . . played a non-negligible role in the formation of the new mechanics. To limit oneself to a fragmentary evocation of these themes, without establishing the general framework from which they are indissociable, would render it all but impossible to characterize in its full scope the mutation of scientific consciousness effected by Galileo in the space of fifty years.

Galileo's science, then—more especially his science of motion—in, as one might say, 'its philosophical setting', is Clavelin's subject; except that to put it this way is to suggest a distinction between the 'science' and the 'philosophical setting' which it is one of the virtues of Clavelin's treatment *not* to make. Rather, every stage of his discussion is concerned to determine *in their functional roles and interrelationships* the components and characteristics of the intellectual framework of the theory of motion. The involvement of this total intellectual framework—of ontological, methodological, epistemological, physical, and mathematical ideas and principles—in the mutations studied by Clavelin, is what constitutes the 'prodigious overthrow' he refers to; and to 'comprehend this in its right dimensions' is to understand not only the difference—the distance—between the two systems, before and after the 'overthrow', but also the path connecting those systems—the route by which the changes could and did occur: not only the changes in an interrelated nexus, but the nexus of changes. We see, therefore, that the phrase 'Galileo's work in its historic singularity' has multiple connotations. It implies attention to Galileo's natural philosophy as a connected whole, in order to see how its principles were understood and used by Galileo ('to place oneself inside the work, to attempt to reconstruct it in its organization and complexity'); in this sense, 'in its historic singularity' means 'as in fact it was'. It also implies attention to the special place of Galileo's work in the course of events of (intellectual) history—where, constituting an 'overthrow', a sharp change of direction, it forms a 'singularity' in an almost mathematical sense. To say this is, of course, to refer not only to the antecedents of Galileo's work, but to its consequences as well: to the question how (in Clavelin's words) 'classical mechanics was able to take shape'. Clavelin is far from identifying Galileo's science with classical mechanics: to understand the former in its historic singularity we must understand also how it differs from the latter. It is clear from Clavelin's statement that he does not draw the conclusion that Galileo's natural philosophy must be considered exclusively 'in its own terms', without reference to the science of a later time. On the contrary, he defines the significance of his own book as lying in this relation of Galileo's science to what followed it, and sees his examination of the work of Galileo as a contribution to our understanding of the development of the mechanics of Newton.

The execution of such a plan requires of the investigator, together with philosophical sophistication, both historical scholarship and mathematical/physical mastery—and both of these with respect not only to technical detail, but with respect also to what might be called 'perspective': the judgment of relationships, of significance, and of likelihoods amongst alternatives in an uncertain situation. In the reviewer's opinion, one reservation has to be made in stating a general commendation of the present work. Its execution displays competence in all the named respects; and in many points it merits high praise. With some frequency, however, the judgments of mathematical and physical perspective appear to suffer from a certain narrowness of technical grasp—especially on the mathematical side—that leads to distortion in these larger-scale assessments. Several examples will be given presently. But it should be most strongly emphasised that this defect does not vitiate Clavelin's analysis of Galileo, if it does diminish its virtue. Clavelin has striven for a very comprehensive view, and has produced a study reliable in its account of details, and reasonable—indeed, the reviewer believes, for the most part correct—in its major conclusions.

To make clear the content of the reservation just expressed, and to provide evidence for it, will require rather detailed discussion of rather restricted points. Such discussion will be offered in the first main section of this review. In the second main section, some issues of wider philosophical bearing will be taken up—no longer in the quasi-judicial spirit of a review, but rather in that of constructive controversy. In this second section the discussion will be less formal; philosophical interpretation and point of view will be more crucially involved; and the reviewer, as there representing his own point of view, will allow himself the use of the first person.

1 The book is in eight chapters, and four main parts. The first part deals with the antecedents of Galileo's natural philosophy, in three chapters devoted, respectively, to the theory of local motion in Aristotle, the tradition of the fourteenth century, and Galileo's own early work. The second part, with a brief introduction and two chapters, discusses the Copernican cosmology and its relationship to Galileo's science of motion. The third part, under the title 'Birth of Classical Mechanics', treats the principles of Galileo's dynamics, as they

developed in his investigations of the motion of heavy bodies; this section, again, comprises an introduction and two chapters. The fourth part, in a single chapter, 'Reason and the Real', is devoted to an examination of Galileo's conception of scientific intelligibility: of explanation or scientific theory, and of the validation of theory. A brief 'Conclusion' follows; and six Appendices discuss a series of special points. There is a chronology of Galileo's career; a bibliography (chiefly of works cited); an index of proper names; and a quite detailed analytical table of contents.

In some ways the first chapter, on Aristotle, is the best in the book. Starting from ontological considerations (the problem of Parmenides and the solution proposed by Aristotle); proceeding through discussion of the organisation of the Aristotelian cosmos, the theory of continuity and the refutation of Zeno; ending with Aristotle's attempts at what Clavelin calls a 'descriptive analysis' of local motion (that is, a set of principles by which the particular instances of local motion can be characterised); the chapter as a whole is a lucid and penetrating commentary upon an important part of the Aristotelian philosophy (it could, for instance, be read with profit in a course on Aristotle, or on the history of philosophy). Clavelin is very far from the vulgar error of treating Aristotle's views as a kind of directly transcribed 'common sense'—or, on the other hand, as grotesque or silly. He is able to present the Aristotelian teaching as a well-articulated structure, intelligible and coherent internally, and intelligibly related both to the problems out of whose analysis it grew, and to the world in general to which it was applied. But he also avoids the pitfall so open to those for whom doctrines are to be considered 'on their own terms': namely, the tendency to treat all criticism of those doctrines as 'external' and therefore based upon misunderstanding—or all changes of doctrine as changes of fashion. Thus, in his discussion of the 'descriptive analysis', he shows clearly the difficulties into which the Aristotelian theory falls on the problem of the 'mover' of a projectile in flight, and the limitations of the conceptual apparatus of the theory when it came to any quantitative characterisation of motions. In short, the treatment as a whole is scholarly and subtle, and, at the same time, thoroughly sane and sensible.

A few passages do, nevertheless, display the weakness that has been referred to above. These occur chiefly in the discussion of the 'continuity' of local motion. Thus we find on pages 53–4 the statement that

> . . . the abandonment of a systematic correspondence between the continuous and number (which expresses itself *inter alia* by the affirmation that two incommensurable magnitudes do not have to one another the ratio of a number to a number [*Elements*, X, prop. 7]) made impossible the deduction, or at least the mathematical definition, of physical magnitudes which nonetheless are essential, such as velocity and acceleration; and prevented the construction, if not of any physical science at all, at least of any numerical physical science.

This statement is indefensible—unless it is taken in a trivial sense. For in the cited proposition of Euclid, 'number' means positive integer (or, perhaps, integer greater than one); and the fact that there are magnitudes whose ratio is not a ratio of integers certainly does not make physical science impossible. Now, it is clearly Clavelin's intention to maintain that *this use of the word 'number'* constitutes an impediment to the science of physics. However, in itself, this use of the word 'number' is only a matter of terminology; in a trivial sense, the usage may be said to make a 'numerical physical science' impossible—but it quite as much makes impossible a 'numerical geometry'. The serious question must be: did the Eudoxian-Euclidean treatment of continuous magnitude prevent the development of a *quantitative* physical science? Since this treatment had the precise effect of establishing the theory of continuous magnitudes (or 'quantities') on a rigorous foundation—after the disruption of that theory by the discovery that not all ratios are ratios of integers—it is hard to see a prima facie case for an affirmative answer.

Now, there was in fact a bar—although not an absolute one—to the development of such notions as those of velocity and acceleration by the Greek mathematicians; but the point is missed by Clavelin. Just before the passage quoted above, Clavelin writes (p. 53) that Greek geometry 'allows, under certain conditions, the construction of proportions between heterogeneous magnitudes'; and he cites Euclid, V, def. 5. This definition explains what it means for A to have to B 'the same ratio' as C to D. The definiens makes it plain that in any such relation, or 'proportion', the magnitudes A and B must be homogeneous, and so must C and D: in other words, a proportion can hold between two pairs of magnitudes that differ in kind (*e.g.* a pair of distances and a pair of times); but each pair taken alone must be homogeneous—and the term 'ratio' is used only to refer to a relation between the elements of such homogeneous pairs. This restriction is explicit in Euclid's preceding two definitions, whose essential effect, expressed in modern terms, is the following: a ratio is determined by any pair of elements of any positive Archimedean ordered semigroup (*i.e.* any pair of elements of any species of positive additive magnitudes). Definitions 5 and 7 then allow the comparison of ratios, with respect to both equality and greater-or-less, not only within the same species of magnitude, but across species; *they do not, however, allow the formation of ratios across species* (in particular, the 'alternation' of a heterogeneous proportion cannot be considered legitimate—or even meaningful—on the basis of the Euclidean definitions). Again in modern terms, the effect is to make ratios equivalent to pure ('dimensionless') real numbers. Such a conception of ratio is of course an elegant—and eminently useful—one; but it does not allow us to speak of density as a ratio of mass to volume, or velocity as a ratio of distance to time.

The bar is by no means absolute; in principle, here too, we have merely a matter of terminology. It is entirely possible, using concepts that were considered legitimate by the Greek geometers, to define the notions of velocity and acceleration (just as Archimedes, for instance, clearly possessed the notion of density—although he does not appear to have given a definition). But although possible, it is far from trivial; it was not in fact done; and to develop such notions in a *systematic* way would precisely have been to develop a systematic notion (under whatever name) of what we should call 'dimensioned ratios'.

In summary, the situation is this: Clavelin dismisses the problem of 'proportions between heterogeneous magnitudes' as no hindrance to a mathematical natural science, but finds such a hindrance in the notion of ratios which are not ratios of numbers. For dealing with such ratios, however, the Greeks had a well-developed theory, which made what they called 'ratios' *equivalent* to what we call 'real numbers' (and what they called 'ratio of a number to a number' equivalent to what we call 'rational number'—it is worth noting that in Euclid's terminology rational fractions also are not 'numbers'). On the other hand, despite their legitimation of 'proportions between heterogeneous magnitudes', of *ratios* of heterogeneous magnitudes—under this or any other name—they had no general theory and no general concept. Any one species of such ratio—density, velocity, *etc.*—constitutes in effect a species of magnitude; these magnitudes could certainly be handled by the methods of the Greeks; but the *general theory* of such ratios is a systematic theory for the introduction of 'new' species of magnitude (*i.e.* the derivation of magnitudes from other magnitudes); and the lack of a theory of this kind constituted a serious limitation upon the fluency not only of expression but of thought about 'dimensioned quantities'.

This issue, although extremely minor in relation to Clavelin's concerns, appeared worth careful discussion both as a symptomatic point and because it has a certain degree of intrinsic interest. Another example occurs a few lines later on page 54: Citing Aristotle's definition of the continuous as 'that which is divisible into parts, themselves always divisible', he characterises this as 'only the formulation, in more philosophical terms, of the axiom of Eudoxus'. This equivalence seems to bespeak a standard of comparison more appropriate to literary criticism of a somewhat impressionistic sort than to the history of science: neither the axiom of Eudoxus nor its most important consequences can be deduced from Aristotle's principle of continuity, for there are 'non-Eudoxian magnitudes' that satisfy Aristotle's condition; moreover, conversely, there are 'Eudoxian magnitudes' (*e.g.* the positive integers!) that fail to satisfy Aristotle's condition—so that the conceptual inequivalence is complete. Again, on page 59, having cited two arguments of Aristotle against Zeno, Clavelin comes to a third, which he says is 'without doubt in Aristotle's eyes the veritable answer to the problem of Zeno'. Part of this quoted argument is fallacious, and indeed has a false conclusion: that no motion with a finite path can take an infinite time, and no motion can exhaust an infinite path in a finite time. The claim is made explicitly 'whether or not the motion is uniform'; but the argument is valid only for uniform motions. Clavelin does not mention the fact that the argument is fallacious; and by referring to it honorifically, as 'doubtless Aristotle's veritable response', and as expressing the essence of the continuum because 'the spring of the argument . . . is just the axiom of Eudoxus', he creates a very misleading impression. (Here, however, the reference to the axiom of Eudoxus is warranted: Aristotle's argument applies that axiom; but—in the case of non-uniform motion—applies it incorrectly. The argument requires the principle: if A is a finite positive magnitude, and $b_1, b_2, \ldots$ a sequence of finite positive magnitudes, then for some n, $b_1 + b_2 + \ldots + b_n$ will exceed A. If $b_1 = b_2 = \ldots$, this is the axiom of Eudoxus. In general, however, it is false; and this very fact is crucial to the refutation of Zeno's 'Achilles' argument.)

Two further mistakes seem to involve small lapses in scholarship. On p. 46, in connection with Aristotle's arguments against the mobility of the earth, we read that 'one always notes that the stars rise and set at the same hour for the same region of the earth'. This is an astronomical error that Aristotle could not have committed; and in the cited place, what Aristotle actually says (according to the version of J. L. Stocks) is that 'the same stars always rise and set in the same parts of the earth': *i.e.* that (*tacitly* presupposing a geographically fixed place of observation) the points of the horizon at which a star rises and sets, unlike those for the planets, are constant. On page 65, Aristotle is said to invoke 'the unhappy example of the cylindrical helix' to illustrate the notion of a trajectory whose parts are not superposable. But it seems gratuitous to attribute this blunder to Aristotle, when his words do not at all explicitly specify a cylindrical helix, referring rather to a mere indefinite 'spiral'.

Turning from these technical criticisms to matters of more philosophical weight, the reviewer has questions to raise on only two or three points; and these are offered only as doubts or alternative suggestions, not as distinct objections to Clavelin's analysis:

(1) It seems questionable whether the term 'motive force' is properly applicable, in Aristotle's theory, to natural motions. In itself, this issue may be more one of words than of principle: for Aristotle does certainly make (quasi-) quantitative statements about the 'impulse of weight or lightness', and one might find it convenient to call such a quantity a 'motive force', even if Aristotle himself did not use a similar phrase. But Clavelin takes the notion of motive force as attributable to the *mover*, the efficient cause, of the motion; and this makes the issue a substantive one, closely connected with the following.

(2) Discussing (pp. 67–8) the problem of identifying the 'mover' of heavy and light bodies, Clavelin concludes that the problem is 'without issue': that Aristotle gives no satisfactory answer, but only two 'indications' on this point—(*i*) that the first cause of such motions is in the heavens, and (*ii*) that the remover of an obstacle is an accidental cause; and that as a 'practical solution', when it comes to evaluating 'the force on which such motions depend', Aristotle simply invokes the weight (or lightness) of the bodies. Now, the problem is without a doubt a vexed one; but it is possible to construe Aristotle's statements somewhat differently. Aristotle does indubitably refer to weight and lightness as internal principles of motion, and does state quantitative relations between these and the motions themselves. But such internal principles or 'natures' seem rather to be considered as *formal* than as 'efficient' causes—i.e. seem to be considered as belonging to the 'essence' or 'substance' of the thing. Accordingly, in treating the problem of the causes of natural motion, Aristotle appeals first to his theory of the process of generation: 'Light is generated from heavy, *e.g.* air from water; for water is the first thing that is potentially light, and air is actually light, and will at once realise its proper activity as such unless something prevents it. The activity of lightness consists in the light thing being . . . high up: when it is in the contrary situation, it is being prevented from rising. . . . But, be it noted, this is the question we are trying to answer—how can we account for the motion of light things and heavy things to their proper situations? The reason for it is that they have a natural tendency . . . towards a certain position: and this constitutes the essence of lightness and heaviness . . . .' (*Physics* VIII 4 255^{b}8-16; *cf. De caelo* IV 3 310^{a}31: 'That which produces upward and downward motion is that which produces weight and lightness'; and what follows, to the end of the chapter.) On this reading, Aristotle's solution to the problem of the efficient cause of 'natural motions' is the following: In the most proper sense, the efficient cause of the *motion* of a heavy or light body, *i.e.* the efficient cause of its actualisation as heavy or light, is just the efficient cause of its *generation* as a heavy or light body; but since a body may be 'actually' such-and-such, and yet prevented from 'realising its proper activity as such', in a secondary sense the remover of an obstacle is efficient cause (*Physics* VIII 4 255^{b}24: 'The thing

in a sense is and in a sense is not moved by one who removes what is obstructing and preventing its motion'). In neither the proper nor the secondary sense, however, does the efficient cause so identified stand in any quantitative relation to the local motion of the heavy or light body.

(3) The alternative reading just suggested by no means invalidates Clavelin's central conclusion in this matter: that Aristotle's 'descriptive analysis' of local motion leads to a serious impasse; but it does change the locus of that impasse, displacing it at least partly into the *ontological* sphere. Aristotle appeals to a distinction of two senses of 'potentiality': water is potentially light as a student is potentially learned; air, in one way actually light, may (*e.g.* in an inverted bottle held under water) be in another way still just potentially light, as (say) a sleeping scholar has knowledge potentially. Then the efficient cause properly speaking is like the teacher; the secondary cause like the alarm clock. However, this analogy leads to an equivocation. Aristotle says that 'air is actually light, *and will at once realise its proper activity as such* unless something prevents it'. He also says that 'the activity of lightness consists in . . . *being* . . . *high up*'. But it is not true that air, 'generated' (*e.g.* by electrolysis!) at the bottom of a vessel of water, *is* 'at once' *high up* (as the wakened scholar may be thought 'at once' to know the Iliad). Aristotle's ontological analysis of motion, as (in effect, as Clavelin rightly puts it) a *process* and not a *state*, enters to resolve this equivocation: what sets in 'at once' is not the full actuality, but the process of actualisation—the 'activity of the potential qua potential'. The trouble is that it is just this intermediate process that Aristotle attempts—however unsuccessfully—to put into mathematical relation with the form, or essence, of the moving body: he relates lightness or weight, not to the height or depth achieved, but to the speed of motion. Moreover, in the analogous relations for violent motions, the principle of motion that plays the role corresponding to that of weight or lightness in natural motions is indeed 'force' (*cf. e.g. De caelo* III 2 $301^{b}17$ ff.), and this is indeed an efficient cause of a local motion (not the form of an activity). So Clavelin's assimilation of weight and 'motive force', whether or not interpretatively sound, in any case corresponds to something in the doctrine—something that on the present construction amounts to a conceptual incoherence in the ontological account of motion.

An allied point can be made with respect to the remaining case of natural motion in Aristotle: the eternal circular motion of the heavenly bodies. Here there is no question of a force: beyond controversy, Aristotle assigns *this* motion to the sphere of activity, essence, and form; it is the full activity of the heavenly substance. But how can this be reconciled with the ontological analysis of motion? How can such a motion be regarded as other than a 'state' of being? Or, to put the problem in its most specific form: how can Aristotle's doctrine on the character of the motions of the heavenly bodies be squared with his ontological principle that change, hence more particularly motion, hence still more particularly local motion, requires for its very definition—that is, for its ontological intelligibility—a *terminus a quo* and a *terminus ad quem*? (Clavelin cites the principle on page 24 [*cf. also* p. 115: 'an initial term and a final term are absolutely required for motion to have a meaning and a function']; but when he discusses circular motion, on page 43, and credits Aristotle with a coherent analysis of its status, he does not raise this question.)

Chapter II, on the conceptions of the fourteenth-century schools of Oxford and Paris, suffers more pervasively from the weakness of technical base that has shown itself in the first chapter only in isolated points. Nevertheless, the chapter (with Appendices I and II, on Oresme's theory of 'configurations' and on 'Bradwardine's law') is richly and reliably informative on the technical notions introduced in that remarkable period; and if, as the reviewer believes, the 'perspectival' judgments of the significance of those technical contributions are wrong in important ways, the objective evidence on which the issues must be decided is presented by Clavelin with thoroughness, care, and clarity. This presentation is an excellent summary and appreciation of the work of the principal cultivators of this field (among whom, of course, the pioneer was Duhem, and the author of the most central and definitive studies has been Anneliese Maier). The longest section deals with the development of fundamental conceptions of kinematics (under the scholastic title of 'analysis of motion in its effects'); it is in this development, as Clavelin remarks, that 'the originality of the fourteenth century manifests itself most brilliantly'. The remaining sections treat the problem of 'moving forces' (the 'impetus' dynamics of Buridan and his disciples), and the fourteenth-century discussions of cosmological questions; and the chapter ends with a summary and appraisal—a 'striking of the balance'—of the contributions of these schools to the theory of motion. This general appraisal is, on the whole, characteristically judicious; its chief points may be summarised as follows:

(1) The most important contributions of the fourteenth-century schools were to the 'descriptive analysis of motion with respect to its effects'—*i.e.* to kinematics: namely, the notion of instantaneous velocity (as an 'intensive magnitude'—a conception to whose development, starting from an idea of Duns Scotus's, Clavelin devotes an interesting discussion); concepts appropriate to the description and classification of non-uniform motions; and, in particular, the law connecting distance traversed with time of traversal for a 'uniformly difform' (*i.e.* uniformly accelerated) motion ('average velocity law'; 'law of odd numbers').

(2) Next in importance (although, according to Clavelin, of far smaller interest for the later development) were the dynamical considerations belonging to Buridan's theory of ***impetus***, which 'contributes to the clarification of the problem of moving forces, by eliminating the role of the medium in violent motions, and by suggesting a less arbitrary explanation of the acceleration of natural motions'.

(3) The development of the 'descriptive analysis of motion with respect to its causes', represented by the (in effect) logarithmic law of Bradwardine connecting velocity and ratio of force to resistance, is by contrast an isolated scholastic episode: Bradwardine's law possessed formal superiorities over Aristotle's, so far as the dialectical discussions were concerned, and it was consequently adopted by all the principal fourteenth-century physicists; but neither from the point of view of the truth about motion, nor from the point of view of the development of classical mechanics, did it have any real significance.

(4) The achievements of the fourteenth century have to be regarded as bounded by the following rather severe limiting factors: (*a*) The frame-

> work of Aristotelian general principles—the ontological principles affecting the conception of motion (motion intelligible only as a 'process', not as a 'state'); the intrinsic physical analysis of motion (relation of mover and moved); the principles of the cosmological order, and the subordination of motion to that order, with the associated distinction of motions into natural and violent and the theory of the elements—all this persists intact and unchallenged. (*b*) The investigations of the fourteenth century, in the spirit of the commentaries, were characterised by a close compartmentalisation of their respective domains. Thus, not only were the results not brought to bear upon a critique of the general principles of motion, but even in points of technical detail there was no attempt to employ new conceptions introduced in a certain context in a discussion that tradition separated from that context. The most striking example is that of falling bodies: no trace is to be found of an attempt to associate the motion of falling bodies to any one of the kinematical classes newly defined and analysed in the fourteenth century—in particular, no suggestion is found that such motion is (or is not!) 'uniformly difform'; and Clavelin notes that 'traditionally the natural motion of heavy bodies belonged to the analysis *quoad causas* and not *quoad effectus*'.

Of these main conclusions, all are defensible, and in the reviewer's own judgment most are entirely correct; only in the appraisal of the value of the theory of *impetus* would the reviewer be inclined to suggest another view. But on a finer scale, the imperfections already mentioned appear; some examples follow, starting with a few rather technical points, and proceeding to broader issues:

(1) In explaining the development of the notion of velocity as an 'intensive magnitude', Clavelin starts from the distinction between processes of change that occur by 'intensification' or 'remission' of a *quality* of the subject, and those that consist in an 'addition' or 'subtraction' of *parts*: the traditional distinction of 'intensive' and 'extensive' magnitudes. He then notes a second distinction, which separates local motion from all others (including growth/diminution, itself a change of an *extensive* magnitude): in all other cases, what changes is an 'internal determination' of the subject, whereas in local motion there is no such determination that changes. To remove this disparity, Clavelin tells us, Duns Scotus introduced the notion of the *ubi*, an instantaneous determination of the mobile which stands in the same relation to local motion as degree of temperature to change of warmth, degree of moistness to change of humidity, or size to growth and diminution. Clavelin then proceeds (pp. 78–9) to remark that this 'comparison' (between local motion and the change of an intensive magnitude) implies a 'reciprocity': 'just as a heating or cooling must be related to the action of these qualities, heat and cold', so must local motion be 'related to the action of a quality'. And this quality can only be velocity: hence velocity as an intensive quantity, characterising a local motion at each instant.

The data in this account are of great interest; but the last argument—which is not attributed by Clavelin to any of the sources, and is presumably his conjectural reconstruction—is, even as dialectics, quite inadmissible (indeed, incoherent). The 'comparison' based upon the *ubi* is already 'reciprocal': what is 'actualised' in local motion, as heat is actualised in heating, is (if anything) *ubi* (*i.e.* 'here or there'), not velocity: indeed, local motion is surely not to be identified with intensification or remission of velocity!

(2) Discussing Oresme's analysis of non-uniform motion with the help of his celebrated *configurationes*, Clavelin refers to the conception introduced by Oresme of the 'quantity of velocity': 'the sum, or at least synthesis, of all the intensities that the velocity has assumed'. In some sense, this notion clearly reaches toward the modern notion of an 'integral' of the velocity; and since Oresme identifies the 'quantity of velocity' with the area of the *configuratio*, and since also he normally chooses for *longitudo* or abscissa of a *configuratio* of velocity a line representing the time, in practice the concept is that of the time-integral of velocity—hence the distance traversed. (Or, as one may also say, Oresme's 'quantitas velocitatis' amounts to the total change of *ubi* in a motion: so that *this* sense of 'velocity' *does* stand to local motion as heat to heating. But Clavelin neglects this point.)

Now, these are exceedingly important considerations; and Clavelin's account (pp. 90–1) makes them quite apparent. But he offers a comment that introduces a very bad confusion:

> In a text justly highlighted by A. L. Maier, we see Oresme implicitly assimilate the *quantity of velocity to the total velocity*, that is to the greatest value attained by the velocity in the course of the motion. But the total velocity, in the Middle Ages, is by definition proportional to the path traversed during a given time: to assimilate the quantity of velocity to the total velocity then amounts quite precisely to considering that the space traversed as result of a process of variation is directly proportional to the surface of the *configuratio* expressing this process.

The last conclusion is very important, and is that of Oresme. The argument, however, contains a double confusion: (*a*) the maximum velocity is (in general) proportional neither to the 'path traversed during a given time' nor to the total space traversed; (*b*) the path traversed during a given time (itself not a well-defined quantity) is (if more precisely defined—say, by specifying the first, or last, or medial, unit of time during the process) not proportional to the total space traversed. To attribute these (or any of these) erroneous proportionalities to 'the Middle Ages' seems rather too sweeping, and not clearly substantiated: and Clavelin does not appear to realise that they are errors. In any case, neither confusion is found in the quoted passage of Oresme. On the contrary, in that passage Oresme concludes that the 'total velocity' of a certain process is four times 'the velocity of its first part'; but during the process in question, the instantaneous velocity *grows without bound*—so *by 'total velocity' Oresme cannot possibly mean maximum velocity*; moreover, in that process a path of length $4s$ is described in two units of time, whereas during its 'first part', of duration one unit, the distance s is covered—so *Oresme is plainly and simply not comparing, under the name 'total velocity', spaces traversed 'during a given time'*. Clavelin's interpretation rests upon nothing in the passage but the words 'total velocity', with the unsupported construction—or pair of constructions—he puts upon them. In view of the tangle of confusions thereby introduced into an otherwise quite lucid analysis, both historical supporting evidence and conceptual analysis are clearly called for; their absence leads one to suspect that the confusions are in fact Clavelin's, and the constructions in question incorrect.

(3) On pages 87–8 and 95–6, we find the discussion, following Swineshead, of a case of 'difformly difform'—that is, non-uniformly accelerated—motion. It is the case of a velocity that changes stepwise, with intervals of constancy forming a geometric progression whose ratio is $\frac{1}{2}$. As to the velocities in successive intervals, they are described on page 87 as being augmented or diminished 'according to a rate fixed arbitrarily', and (more particularly) on page 95 as 'doubled successively in each proportional part of the time'. Yet in both places the sequence of velocities is in fact given as the arithmetic progression a, $2a$, $3a$, $4a$, . . . , na, . . . ! Further, we are told on page 95 that 'the fourteenth century knew this mode of variation under yet another form': in this 'second form', the successive velocities are the same as before, and the corresponding time-intervals —instead of 1/2, 1/4, 1/8, *etc.*—are 1, 1/2, 1/4, *etc.* Since this 'difference of form' consists only in a choice of a new unit of time, it is hard to see why it should be worth mentioning. Doubtless it is a correct scholarly remark that these two modes of description appear in the texts; but to make a point of this is surely to over-emphasise purely literary considerations; and to say that 'the fourteenth century knew' these *two* 'forms' of the process is to underestimate badly the mathematical understanding of the thinkers concerned—by suggesting that they did not 'know' equally well the 'forms' in which the times are called b, $b/2$, $b/4$, *etc.*, with *any* given b. (In point of fact, Oresme, in the passage referred to in (2) above, describing just this mode of variation, speaks of a mobile which 'is moved, in the first proportional part *of any time* so divided, with any velocity; and in the second is moved with double velocity; and in the third with triple [*etc.*]'.) Finally, in the same connection, Clavelin remarks (60) that 'this mode of variation produces the same results as a uniform variation'. He presumably refers to the fact that the distances traversed in the first half and the second half of the total duration of this process have the ratio 1:3. But this is a very superficial (and very limited) coincidence; the properties, or 'results', of this motion are clearly altogether different from those of uniform acceleration: *e.g.* in this motion the speed becomes infinite in a finite time! (Two glosses on this passage: (i) Clearly a motion which 'produces the same results' as a uniformly accelerated motion must *be* a uniformly accelerated motion. (ii) A motion uniformly accelerated from rest at time t describes distances whose ratio is 1:3 in *every* pair of equal time-intervals of which the first begins at t and the second immediately follows the first. The motion in question does not possess this property. Moreover, even this stronger property is not sufficient to characterise uniformly accelerated motion.)

(4) In discussing the arguments by which the 'mean-velocity theorem' for uniformly accelerated motion was established by the scholars of Merton College, Clavelin remarks that the proofs were numerous, and proceeds (p. 93): 'We shall leave aside the purely syllogistic proofs, much in favour at Oxford, and restrict ourselves to those that rest explicitly upon a quantification of the process of variation'. The example chosen for discussion is a very interesting demonstration by Swineshead. However, it is *prima facie* most puzzling that any proof of such a result could have been 'purely syllogistic', and have failed to rest upon a 'quantification of the process of variation'. The example cited by Clavelin of a proof of this sort (footnote 52) is one given by Heytesbury. When one checks the cited text, one finds that the form of Heytesbury's argument is indeed syllogistic; but

that (of course!) the ultimate premises of the series of syllogisms are propositions, taken as evident, about quantitative characteristics of uniformly accelerated motions. What is altogether striking, however, is the fact that the basic idea of Heytesbury's proof is *precisely the same* as the basic idea of Swineshead's. To distinguish between the two in the terms already quoted, suggesting a rather fundamental contrast, and not to record the identity of the quite remarkable key principle of both, is in the reviewer's opinion a regrettable subordination of conceptual substance to merely rhetorical form.

(5) A somewhat similar tendency is seen in the following case, also involving a comparison of Heytesbury and Swineshead, in which the issues involved are of considerably wider significance. We have already seen the general assessment that Clavelin makes of the fourteenth-century contributions to the theory of motion, and of their limitations. In respect of the contributions to kinematics—which he ranks most highly—he offers a series of more special judgments of a rather negative character. One of these is the claim that the kinematics of the fourteenth century was not 'logically articulated' as 'an ordered whole.' Now, this claim may, in some sense and to some degree, be justified and important; but in what sense, and to what degree, remains obscure; in any event the following passage (p. 97) is wide of the mark:

> Certain texts might cause one to believe that the 'law of distances' is articulated logically upon the 'law of the mean degree', as Galileo will have it, and as is necessary if one is to have a coherent science of motion. In fact, this is so little the case that one sees the two propositions deduced alternatively the one from the other; Heytesbury, for example, deduces the law of distances from the law of the mean degree; Calculator, on the contrary, bases himself expressly upon the law of distances to prove the law of the mean degree, showing that the second would be violated if the first were not true [*sic*; but evidently 'second' and 'first' should be exchanged].

In the first place, the statement that 'Calculator'—that is, Swineshead—'bases himself upon' the law of distances to prove the law of the mean velocity is somewhat misleading. Swineshead gives a series of proofs of the mean speed theorem; one of these, as has already been remarked, is essentially the same as that of Heytesbury. It is true that *one* of Swineshead's proofs derives the mean speed theorem from the law of distances; but the proof in question is not 'based upon' the latter, since the distance law itself is established in the course of the argument: that is, Swineshead (by a most interesting piece of reasoning) establishes the law of distances directly, and then derives the law of the mean speed from it. Clavelin's words, implying a defect of 'logical articulation' and a lack of 'coherence', suggest—if they do not strictly imply—a net circularity in the reasoning of Heytesbury and Swineshead taken together. No such circularity is present.

But what is more seriously wrong with this passage is its underlying premise, that there is one special order of 'logical articulation' of propositions or laws that must necessarily be followed for a science to be 'coherent'. This notion seems to reflect a quite serious misunderstanding of the methodology both of mathematics and of empirical science. Swineshead's interest in alternative proofs, by different routes, involving (in part) opposite sequences of conditional 'dependence', may be an expression of the spirit of scholasticism; yet it is no less a characteristic interest of the sophisticated mathematical investigator, and would appear anything but bizarre to a mathematician or theoretical physicist of our own day.

The worst consequence of this assumption of a canonically correct order of exposition—or of deduction—is that, there being in fact no conceptual basis determining such an order, the assumption can only lead to the adoption, as standard, of the order that is found in some 'classical' modern text. In other words, as in the preceding case (4), it becomes a matter of rhetorical form, or expository style, taking on the role of criterion for judgment that ought rather to turn upon conceptual content: a substitution of an essentially literary standard for a scientific or philosophical one.

(6) Clavelin argues that no clear concept of *acceleration* was possessed by the fourteenth-century schoolmen. The assertion is first made about Oresme in particular (p. 83): 'A considerable distance no doubt separates the *velocitatio* of Oresme and acceleration in the modern sense; Oresme does not know the idea of an instantaneous variation of velocity, and the *velocitatio* only designates the manner in which the augmentation of velocity operates in the whole time of motion'. The evidence upon which this assertion rests is presented on pages 99–103. This evidence is, in itself, of considerable interest: it bears upon the question of the ontological interpretation of the processes of 'intension' and 'remission' of qualities, or of 'intensive quantities'. The Scotists, and William of Ockham, viewed such processes as occurring *by addition and subtraction of increments*; that is to say, they regarded the *differences of intensive magnitude* as, themselves, truly 'individuated' magnitudes, which are adjoined to or detached from the initial magnitudes. Clavelin argues that the texts indicate that this view was *not* shared by the physicists of Oxford and Paris. His argument, in general, is far from conclusive; but in the case of Oresme he is able to cite a decisive passage (p. 103): 'There is [in the intension of any intensible quality] neither plurality, nor real accumulation of degrees, as certain people think. The quality, during the whole time of alteration, is a *successive* quality, different in every part of that time'. In particular, then, for Oresme, a changing velocity is a *wholly new velocity at each instant.* This conclusion—in which Clavelin agrees with Anneliese Maier against Duhem and Clagett—is unquestionably sound.

The problem lies in the inference Clavelin makes from this conclusion, supposed as valid for all the physicists in question (p. 103; italics added here):

> There is not, *therefore*, in the fourteenth century, any quantification of the concept of acceleration, in the strict sense. One could not even say, without abuse of terms, that the idea of acceleration is present in the Mertonians and the Parisians. They saw indeed that a velocity may be affected with diverse variations, and they described these; but the variations always remained total processes, distributed over the whole time of the motion; they never conceived *the idea of a variation in an instant*, *by addition of a determinate quantity of velocity.*

This 'idea', as here formulated (and described on p. 102 as 'one of the capital acquisitions of Galileo'), is, taken strictly, quite incompatible with the concept of acceleration: addition of a determinate, non-zero, increment of velocity in an instant characterises a discontinuously varying motion; whereas the idea of a determinate 'infinitesimal' increment, which is what Clavelin doubtless intends, is not a clear and well-defined conception at all. On the other hand, if one turns to the passage from Oresme cited by Clavelin (p. 82) immediately before his first denial to Oresme of the concept of acceleration 'in the modern sense', one finds that Oresme says the following about *velocitatio*:

> Every velocity can be increased or diminished. Its continuous increase is called *velocitatio*, and this *velocitatio* or augmentation of velocity can itself be quicker or slower. Whence the fact that sometimes the velocity is increasing and the *velocitatio* is decreasing, whereas at other times both increase together.

This passage—especially the last sentence—is extraordinarily hard to reconcile with Clavelin's commentary upon it. To describe the *velocitatio* as itself varying in a definite sense over a time-interval, and to distinguish so emphatically *its* variation from that of the velocity, is surely *not* to regard the *velocitatio*, qua variation of velocity, as a 'total process, distributed over the whole time of motion'; but rather to regard it as an 'intensive quality' in its own right—indeed, for Oresme, to regard it, like velocity, as a 'successive' quality, 'different in every part of that time'.

If the passage from Oresme allows any margin of doubt as to the strict parallelism, for him, of the quantitative status of velocity and of acceleration, a quotation from Swineshead given by Clavelin on page 87 allows no such margin at all: Swineshead affirms explicitly that '*intensio motus* is related to motion in the same way as motion to space: for as space is traversed by means of motion, so motion is acquired by means of *intensio motus*'. Swineshead proceeds (this is not quoted by Clavelin) to explain the quantitative analogy of the *measures* of motion and of *intensio motus*: the former is measured by the line described in a given time, the latter by the 'latitude of motion' acquired—*i.e.* the increment of velocity—in a given time. It is worth noting that this strict analogy would make it altogether natural to proceed to further—indeed, to arbitrary—time-derivatives of a variable: the conception that underlies the analogy, as Swineshead expressly puts it, is that an 'intensive quality' is 'acquired by means of' another such quality, its 'intensio'; and this principle clearly serves to generate an unending sequence of derived 'qualities'. Although the fourteenth-century physicists may not have gone so far as to generalise their conception in this way, the mere fact that they possessed a conception so immediately susceptible of such generalisation is testimony to the 'modern sense' in which they understood acceleration—to which the (intrinsically interesting) ontological issue discussed by Clavelin is irrelevant.

Before leaving the consideration of the fourteenth century, one small but rather striking point is worth remarking upon. On pages 115–18, Clavelin considers certain metaphysical analyses of the problem of motion, space, and the void. In the course of the discussion, he refers to Duhem as having attributed to Oresme a conception of 'absolute space' that anticipated the conception of Newton. Clavelin rejects this interpretation of Oresme's view, on the grounds that Oresme did not, like Newton, conceive motion as a 'state' of the moving body; in particular, a state with a tendency to persist. (It may, however, be questioned whether the concept of motion as a 'state' and the concept of 'absolute space' have so intimate a connection; at any rate, such contemporaries of Newton as Huygens and Leibniz accepted the former and rejected the latter—and the reverse would seem to be even easier.) Clavelin then remarks (p. 117) that the whole discussion, in Oresme, takes place in a theological context:

> Some pages earlier, Oresme speaks of space as of 'the immensity of God himself', repeating almost word for word an opinion maintained by Bradwardine, who saw in space the infinite place of the divine power.

The striking point—ignored by Clavelin—is that this theologico-physical position of Bradwardine and Oresme (on which *cf.* further Clavelin's citations on p. 118) is very close indeed to the view of Henry More and of Newton himself on the relation of space to God. It is so close that the question of a possible continued tradition, of an actual influence, quite naturally arises. In any case, we certainly do not have here a basis for denying an anticipation of Newton by Oresme!

2 This discussion of the early part of Clavelin's book should suffice to give an idea of its strengths and weaknesses. It is clear that a treatment of the entire work on the same scale would produce a commentary rather than a review. The remainder of this essay will be devoted, instead, to the consideration of two large issues in the philosophical interpretation of Galileo. In differing with Clavelin on these issues, I shall not (as in previous cases) be expressing the judgment that his treatment of these issues is technically deficient (although, on details, criticisms of that nature may of course occur). Rather, it is a merit in a major philosophical study of a major figure to provide a focus for serious discussion of important controversial questions; and it is in the spirit of such constructive controversy—and in explicit recognition of the merits of this work—that I proceed to the following matters.

(*a*) *The relation of Galileo's Copernicanism to his mechanics*

In the introduction to his Second Part: Copernicanism and the Science of Motion, Clavelin writes as follows (pp. 181–2):

> From 1602 to 1642 the researches of Galileo will arrange themselves about two essential preoccupations: the justification of Copernicanism on the one hand, the construction of a mathematicised science of the motion of heavy bodies on the other. Two works will mark their outcome: the *Dialogue on the two principal systems of the World* in 1632, the *Discourses and mathematical demonstrations concerning two new sciences* in 1638. . . .
>
> Now it is easy to show that in the formation of classical mechanics a role of the first rank, hardly less important than that of the *Discourses*, is to be attributed to the Galilean cosmology. Since antiquity, in fact, heliocentrism had encountered a difficulty of a general, and as it were prejudicial, order: how to reconcile the possible movement of the Earth, and especially the diurnal motion, with ordinary mechanical experience. . . . In order to remove this objection—or rather the series of objections to which this difficulty gives rise—Galileo inaugurates an entirely original approach: enunciate the traditional arguments against the diurnal motion in the form of mechanical problems; then show that, so specified and interpreted, they

> lose all force, certain principles of the science of motion allowing in each case the reconciliation of the rotation of the Earth with everyday mechanical experience.

This seems to me, so far, a thoroughly correct and admirably clear statement of the case. But Clavelin proceeds (pp. 182–183):

> In fact the true problem is to evaluate that contribution [of cosmology to mechanics] at its true rate, and especially compared to that of the *Discourses*. The approach adopted by Galileo suggests a solution. To remove the arguments against the diurnal motion, Galileo . . . uses principles and concepts taken from the science of motion: everything therefore lets one suppose, at first view, that the mechanical analyses of the cosmological work apply to special cases a science previously elaborated—just that which the *Discourses* will later divulge. . . .
>
> Without being radically incorrect, such a reconstruction nevertheless has serious defects. The most important, without a doubt, is that it tends to efface the differences (although they are evident) that separate the *Dialogue* and the *Discourses*. It is in fact striking to note that . . . the *Dialogue* makes no direct use of the very precise theorems and propositions that constitute the science of naturally accelerated motion and of the motion of projectiles. . . . But the principles used in the one and the other work present a no less pronounced contrast. It is remarkable, for example, that one of the principles whose role is decisive in the *Dialogue* [and here a footnote names the principle of mechanical relativity] is not even enunciated in the *Discourses*; as to the principles that figure in both works, neither their formulation nor the manner in which they intervene in the reasoning are really comparable. Now it suffices to consider, even briefly, the nature of the problems debated in the *Dialogue* and in the *Discourses*, to understand that, far from being accidental, these divergences rather illustrate the characteristic irreducibility of the two investigations.

Here I disagree, both on the conceptual and on the literary issues. As to the latter, in my view the difference in logical organisation, in formulation, and in argumentation, between the *Dialogue* and the *Discourses*, has far less to do with a radical distinction in the investigations they concern whan with a difference in rhetorical intent: the earlier book was addressed to a wider audience than the later one, and treated matters of great—even sensational—popular interest. Moreover, quite apart from their cause or occasion, the facts themselves of the literary relationship between these books are exaggerated in the quoted passage. For instance, in Day Two of the *Dialogue*, Salviati states: 'The acceleration of straight motion in heavy bodies proceeds according to the odd numbers beginning from one. . . . Or we may say that the spaces passed over [starting from rest] are to each other as the squares of the times'. And when Sagredo asks whether there is a 'mathematical proof' of this statement, Salviati replies: 'Most purely mathematical; and not only of this, but of many other beautiful properties belonging to natural motions and to projectiles also, all of which have been discovered and proved by our friend'. How, then, can it be maintained that 'the *Dialogue* makes no direct use of the very precise theorems and propositions that constitute the science of naturally accelerated motion and of the motion of projectiles'? An issue might be made of whether the theorem quoted (as well as others) are 'used', or only cited, in the *Dialogue*; but such an argument would skate on rather thin ice. At any rate, one can scarcely doubt that the *Dialogue* does indeed—as a literary work—have available as a resource 'a science previously elaborated—just that which the *Discourses* will later divulge'. Even more: when Salviati goes on to say that he has 'seen and studied' all these properties, which had been 'discovered and proved by our friend', it is hard to escape the conclusion that the Latin treatise *De motu locali*, read and discussed by the participants in the *Discourses* on Days Three and Four, had even been *written down* earlier than the *Dialogue*.

But the deeper question is that of the conceptual and methodological connections of Galileo's cosmological and his mechanical researches. And on this point, I believe that Clavelin is quite right to dismiss as over-simple the view that Galileo *first* elaborated a science of motion, and *then* 'applied' this science to remove the traditional objections to a motion of the earth. However, the true relationship of the two investigations appears to me to be, not the radical divergence or 'characteristic irreducibility' that Clavelin finds, but a closer and more intricate dialectical interplay than is postulated by the naive view that we both oppose.

Galileo's characteristic mode of theoretical procedure is *the elaboration of principles in the treatment of concrete problems.* Thus, unlike the fourteenth-century physicists, Galileo did not content himself with handling special problems in isolated compartments, but had always an eye out for unforeseen relevance; that this was a definite and deliberate policy with him is abundantly clear from the conversation of his dialogical characters. And unlike, for instance, Descartes on the other hand, Galileo did not attempt to formulate with precision the most general principles of the science taking shape at his hands—presumably because those principles had not reached a degree of precision in his own mind that altogether satisfied him (although other reasons too may have operated here).

Now, as Clavelin has remarked in the paragraph quoted at the beginning of this section, Copernicanism set for Galileo a whole series of dynamical problems —which, in another perspective, were also parts of the *general* problem of a radical revision of the physics of local motion (a revision that the new cosmology obviously demanded). On the other hand, even without the shift in cosmology, the traditional physics of the motion of falling/floating bodies and of projectiles had well-known and serious difficulties, to which the fourteenth-century discussions of Bradwardine, Buridan, and others attest—and which those discussions clearly did not resolve.

In my view of the work of Galileo, it is perhaps his most profound achievement to have made, in this situation, a virtue of his necessity: to have found, precisely in reflecting upon the cosmological difficulties, one of the fundamental clues to the solution of the problems of 'everyday mechanical experience'—Clavelin's expression—that the traditional physics had never been able to deal with satisfactorily. In doing this, Galileo (as he proudly asserts) created 'a very new science dealing with a very ancient subject'; he also produced a powerful dialectical support for the new cosmology he had embraced. (But the curve of this dialectical process does not come to its full terminus in the work of Galileo. The dynamical principles of Galileo, refined and developed by Huygens, led finally to the universal mechanics of Newton. Newton offers his mechanical concepts and laws as the essential results of his predecessors, 'confirmed by abundance of experiments'; and, applying these mechanical principles to motions in the heavens, he shows that, of the current cosmological hypotheses, only the Copernican is—approximately—compatible with them. The whole series of reflections, starting from the difficulties posed by Copernicanism, ends thus by establishing the latter as the only viable position.)

The fundamental clue discovered by Galileo in the cosmological situation was the principle of dynamical relativity. For the central difficulty was: how can the earth's motion not give rise to perceptible disturbances among bodies on the earth? And the principle of dynamical relativity is nothing but the reformulation of this point, not as a 'difficulty' but as a general rule: a certain general class of motions, to membership in which the motion of the earth approximates, has the property of *never* affecting the behaviour of bodies which (besides their several particular motions) share a common motion of that class.

That Galileo did possess such a principle of dynamical relativity, and did, in the *Dialogue*, both argue for it, and use it to solve objections to Copernicanism, is well known and uncontroversial. The principle as formulated above is vague, in so far as it fails to specify precisely the class of motions concerned; and Galileo himself never offers a definitive declaration on the point. But it is clear that, if one assumes *rest* to persist in the absence of any perturbing factor, then any motion of the class in question must equally tend to persist: the 'law of inertia' for such motions is an immediate consequence of the principle of relativity. The difficulty of characterising the distinguished class of motions leads to a corresponding difficulty in the law of inertia. The earth's motion, to which the principle of relativity is supposed to apply, is (roughly) 'circular'; and we have, correspondingly, Galileo's principle of 'circular inertia'. Yet it is quite impossible to maintain coherently that uniform circular motion is the motion to which dynamical relativity applies, and it is very likely that Galileo was more or less aware of this. (Two pieces of evidence: (1) Galileo's theory of the tides, which attributes a perceptible effect to the earth's motion; (2) the fact that in the *Discourses* the principle of conservation is applied to motions *in arbitrary directions* at the earth's surface. On the latter point, see the scholium following Proposition XXIII on naturally accelerated motion [pp. 242–4 of the *Edizione nazionale*], where ascending motion on an inclined plane is treated as a composite of 'two different states': one a motion whose speed 'is permanently maintained during the ascent', so that it 'would if acting alone carry the body at a uniform rate to infinity'; the second a uniformly accelerated motion of descent along the inclined plane, which, when superposed upon the former, eventually cancels it, so that the rise comes to an end and a fall supervenes.)

In any case, whether or not Galileo explicitly realised the difficulty that remained in his principle of relativity, the close connection between that principle and his conception of inertia is very clear from the fact that the same difficulty does present itself equally in both. It is also plain that the concept of the 'independent composition' of an inertial and an accelerated motion is a corollary of dynamical relativity. Therefore, although it is quite true that relativity is not directly invoked in the *Discourses*, it seems to me wide of the mark on Clavelin's part to take this as exhibiting a radical 'irreducibility' of the *Discourses* and the *Dialogue*. More characteristic, in my view, of the relationship of the two works is the circumstance that Galileo's solution of the famous objection about the stone dropped from the mast of a ship in motion—in Day Two of the *Dialogue*—is precisely equivalent (despite differences of expository form) to the theory of projectile motion in Day Four of the *Discourses*. This is the example *par excellence* of what I have called Galileo's 'making a virtue of his necessity': for the explanation of projectile motion had been a problem at least since Aristotle; and Galileo succeeds, on the basis afforded by the imagined experiment of the ship—corresponding (by analogy, on the Copernican hypothesis), or *reducible* (by the principle of dynamical relativity), to the unproblematic case of a stone dropped from a tower—not only in giving a full account of the dynamical factors involved in such motion, but even—once his theory of falling bodies is developed—in giving a *complete mathematical characterisation* of the motion: a complete 'descriptive analysis', to put it in a quite un-Galilean way, *both* 'quoad causas' *and* 'quoad effectus'.

To conclude discussion of this point, I should like to suggest a conjectural explanation for the striking omission of any reference, in the *Discourses*, to the principle of dynamical relativity. My suggestion is that Galileo was concerned to avoid, in this work written under the shadow of the Church's ban, any overt reference to the condemned cosmological principles; and that the principle of relativity was too closely identified with Copernicanism to appear safe in this respect. Consider the words of Salviati, declining public discussion of a cosmological point on Galileo's behalf (Day Four, just before Proposition Four): 'He did not wish to speak of it, lest in view of the odium which his many discoveries had already brought upon him, this might be adding new fuel to the fire. But if anyone desires such information he can obtain it for himself from the theory set forth in the present treatment'. Although the point referred to here is not the principle of relativity, the words are quite applicable to it: one can infer from Galileo's theory of projectiles in Day Four, for instance, that the fall of a stone along the mast will appear the same on a ship moving uniformly as on one at rest. Moreover, Salviati's hint—that *the theory of projectiles has bearings upon the questions of cosmology, there for all to see*—is pretty broad; and Galileo could surely count on any competent reader's appreciating the relationship of the theory of Days Three and Four of the *Discorsi* to the discussion in Day Two of the *Dialogo*.

(*b*) *The relationship of Galileo to Plato*

In the course of his extended discussion, in Chapter VIII, of Galileo's conception of scientific explanation and of the roles of reason and experiment in the validation of scientific theories, Clavelin addresses himself to the well-known thesis of Koyré: that for Galileo, as a follower of Plato, 'it is unnecessary to have recourse to experiment in order to recognise the truth'; that, rather, 'good physics is done *a priori*'.

Clavelin's treatment of this issue exhibits his characteristic intellectual balance and good sense. Giving full credit to Koyré (and Tannery before him) for pointing out how far both Galileo's methodological doctrine and his actual procedure were from simple inductivist empiricism, he firmly rejects their extreme conclusions about the apriorism of that doctrine and about Galileo's neglect of experiment in practice. In particular, on the latter point, he finds it impossible (p. 433) 'to subscribe unreservedly to the criticisms that Tannery and

Koyré agree in directing against the experiment alleged, in the third day of the *Discourses*, to validate the law of the square of the times'. The grounds he presents for his assessment are excellent; the assessment itself appears quite sound.

And yet, at a crucial point, a curious ambiguity of judgment occurs. In his final summing-up (p. 465), Clavelin tells us (justly in my opinion) that the study of Galileo 'allows us to perceive to what extent modern science, far from being the negation of philosophy, rests in its origin ['*en son principe*'] on certain properly philosophical proceedings'; and he concludes:

> In placing us at the very sources of [classical] physics, Galileo recalls to us that, born of the substitution of one explicative ideal for another, [that physics] is also a wager which, setting aside its successes, is one of the most beautiful ventures of reason. Against the ever-recurring temptation of positivism or pragmatism, there is no better guarantee than a return to Galileo.

We may pass over the stricture on positivism and pragmatism (which have not been subjected to analytical appraisal in the book, and which I should not myself wish to condemn—or to praise—without first making certain distinctions); the important point is the general one: that significant philosophical illumination for our understanding of modern physics is to be obtained from Galileo. What seems to me an ambiguity of judgment is the contrast between this passage—which from its position in the book, its decisive (even lapidary) formulation, and its relation to Clavelin's undertaking in its entirety, may be presumed to be its author's deeper conviction—and an earlier statement, made in the midst of the discussion of Tannery and Koyré. Speaking of Galileo's method, he there says (p. 434):

> As erroneous as it would be to underestimate the place occupied in it by the examination of experience, it would be equally erroneous without more ado to connect that place with the explicative procedure of modern science. No theorem of the *Discourses*—this is certain—is induced from experience; all are the product of an investigation whose principles have undoubtedly been pondered over in connection with facts, but in which the part of reason remains largely predominant.

The disconnection from 'the explicative procedure of modern science' seems here too lightly assumed.

It is because I share what I take to be Clavelin's own true conviction, that the natural philosophy of Galileo—and especially his philosophy of the character and the method of rational explanation—is of living interest for the philosophy of science, that I consider it worthwhile to press the issue of Galileo and Plato. But of course a thorough discussion of this issue must turn critically upon the philosophical interpretation not only of Galileo, but also of Plato; and is thus a venture far beyond the scope of the present essay. I shall restrict myself here to a few brief comments, intended to open issues rather than to settle them.

(1) The interpretation of Plato is not brought under serious consideration by Clavelin. In contrast to his thoughtful, scholarly, and subtle analysis of the physics of Aristotle, his treatment of Plato is essentially stereotyped. Thus, we

find as rubrics for the discussion of Galileo's alleged Platonism—Table of Contents, page 504—the following negative judgments: 'the principles on which explanation is based are for Galileo not innate cognitions'; 'his conception of scientific truth is not apriorist'. These judgments I consider perfectly sound; and if sound they clearly defeat the attribution to Galileo of Platonism *in the sense of Koyré*; but to say this is by no means to settle the question of the relationship of Galileo's natural philosophy to Plato.

(2) Galileo's not infrequent declarations in favour of Plato are explained by Clavelin as a tactic intended to ward off accusations of heterodoxy, by placing his views under the protection of a respected tradition. The motive is plausible; but it could not be sufficient, if the gap between Galileo's philosophy and Plato's were—to Galileo himself—anything like what Clavelin claims (pp. 462–3):

> From Platonism Galileo asks an alibi against the peripatetic tradition: he does not take from it any concept, any methodological idea; the explanatory ideal that presides over the *Dialogue* and the *Discourses* is as foreign from the speculations of the *Timaeus* as from those of the *Physics* or the *Treatise on the Heaven.*

If this were so, Galileo's literary invocation of Plato would be doubly misguided: as a protective device, it would altogether lack credibility; and from a philosophic point of view, it would be not only unprincipled, but calculated to obscure the very ideas that Galileo's works were intended to propagate. Surely Galileo not only was not, but did not wish to be taken for, an apriorist; surely he did not mean his cosmology to be regarded as comparable to that of the *Timaeus.* Therefore the sense in which Galileo presented himself as allied to Plato must have been different from the sense in which Clavelin rejects the claim.

(3) A possible literary clue to the relationship of Galileo to Plato is quite neglected by Clavelin: namely, Galileo's use of the dialogue form. The absence of any discussion of the significance of this rhetorical mode seems to me unfortunate; for I believe that analysis of Galileo's dialogical exposition can be rather rewarding—and especially so in connection with the *Discourses* (despite the fact that this work is generally considered somewhat dry and perfunctory in literary construction, in contrast to the more scintillating *Dialogue*). It is possible, for instance, to read the opening pages of Day One of the *Discourses* as a little treatise on epistemology—if one is faithful to the well-known principle of Platonic interpretation, that what is displayed in the situation and action of the dialogue deserves as much attention as the direct utterances of the protagonist. The discussion opens in the Venetian arsenal, with a reflection by Salviati upon the relation of the inquiring mind to the practical activity of the artisans. We are at once presented by Sagredo with a problem, in the form of a seeming paradox—more precisely, a phenomenon, reported by the expert workers, which Sagredo finds inexplicable, and whose explanation by the workers themselves (namely, that in mechanics 'one cannot argue from the small to the large, because many devices which succeed on a small scale do not work on a large scale') seems to him irreconcilable with the foundations of the science. The problem itself eventually (in Day Two) becomes the subject of the first of Galileo's 'two new sciences'. The artisans' empirical report was quite correct—but not, of course, their explanation; however, the true relationships, when Salviati first hints at them, are even more bewildering to Sagredo than the inadequate formula of the

workers. The developed theory, of course, is clear and convincing: not only does it remove the mystery (as Salviati foreshadows, on the fifth page, when he says, although in a much more restricted connection: 'Surely this is the explanation and the moment the cause is known our surprise vanishes . . .'), it reveals a significant body of previously unsuspected facts. But this theory is not presented at once; instead, the whole First Day is devoted to a wide-ranging discussion, full of (perhaps somewhat self-consciously Platonic—or Socratic) digressions, which manages eventually to introduce the problem of falling bodies as well—that is, the subject of the second 'new science'. A large part of the material of these preliminary 'Discourses' is incompletely digested, and represents Galileo's tentative thoughts on subjects that he had not—and knew he had not—brought to satisfactory conclusion. The movement of the whole work is from 'Discourses', which represent inquiry, to 'Mathematical Demonstrations', which become possible where science has been achieved. The notion of 'inquiry' suggested by Galileo's exposition is one which it is not far-fetched at all to compare with Plato's notion of 'dialectic'; and the movement, through inquiry, from fallacies and paradoxes to science, invites comparison (for instance) with the 'divided line' of the *Republic.*

(4) I mean to suggest, of course, not only that the historical question of affiliation of Galileo with Plato deserves to be discussed in the context of considerations like the preceding, but that this wider epistemological/methodological context is the right one for the philosophical appreciation of Galileo's own work. Brief reference to Descartes may help to illustrate the point. Both Descartes and Galileo professed the ideal of a mathematical and (in some sense) 'rational' physics. In the detailed work of each one finds a fair share of fantastic ideas and of faulty demonstrations. Yet in the work of Galileo, there is a powerful tendency for mistakes to be corrected; and in his mature exposition, half-baked notions are quite clearly segregated from truly ripened conceptions. Descartes, on the other hand, almost never corrects a mistake; and his definitive exposition of his physics contains a hopeless intermixture of suggestive ideas and quite grotesque ones—with all pronouncements made uniformly *ex cathedra.* Clavelin is certainly right to maintain that 'apriorism' of the sort thus exhibited in practice by Descartes is eschewed, both in practice and in principle, by Galileo. My suggestion is that Galileo's own view of his subtle, flexible, and discriminating method has in common with Plato two things: First, the reconciliation of a 'mathematical' and 'rational' conception of science with a full recognition of the difficulty of coming to know: a reconciliation that is only possible through a crucial emphasis upon the process of inquiry itself. Second, a conviction that inquiry as such is not a profitable subject for positive doctrinal exposition ('Discourse on Method', 'Rules for the Conduct of the Understanding'), but has to be learned and taught through its practice—a process which, in turn, may be foreshadowed in dramatic representation. In the *Two New Sciences*, on this reading, Galileo has offered us his view of *what science is*, in the form of a dramatised commentary upon the central work of his own scientific life.*

HOWARD STEIN
Columbia University

---

* [Editors' note]: Since this review went to press an English translation of Clavelin's work has appeared called *The Natural Philosophy of Galileo*, M.I.T. Press, $25.00.

Brit. J. Phil. Sci. 25 (1974), 398–400 *Printed in Great Britain*

Recent Publications

BOOKS RECEIVED

Agar, M. [1974]: *Ethnography and Cognition.* Minnesota: Burgess Publishing Company. $0.95. Pp. 32.

Arp, H., Field, G. B. and Bahcall, J. N. [1973]: *The Redshift Controversy.* Massachusetts: Addison-Wesley/W. A. Benjamin Inc. $19.50 hardback, $11.00 paperback. Pp. xv+324.

Audi, M. [1974]: *The Interpretation of Quantum Mechanics.* Chicago and London: The University of Chicago Press. £5.75. Pp. ix+200.

Bambrough, R. [1974]: *Wisdom: Twelve Essays.* Oxford: Basil Blackwell. £4.50. Pp. viii+300.

Barnes, H. [1974]: *Sartre.* London: Quartet. £2.75 hardback, £1.50 paperback.

Belinfante, F. J. [1973]: *A Survey of Hidden-Variables Theories.* Oxford: Pergamon. £8.00. Pp. xix+354.

Berkson, W. [1974]: *Fields of Force.* London: Routledge & Kegan Paul. £6.50. Pp. xiii+370.

Botha, R. P. [1974]: *The Justification of Linguistic Hypotheses.* The Hague: Mouton & Co. D.Fl. 70. Pp. 350.

Casteneda, H.-N. [1974]: *The Structure of Morality.* Springfield, Illinois: Charles C. Thomas. $12.75. Pp. 239.

Chattopadhyaya, D. P. [1973]: *Societies and Cultures.* Bombay: Bhavan's Book University. Rs. 6.00 hardback, Rs. 4.00 paperback. Pp. vi+151.

Cohen, R. S. and Wartofsky, M. W. [1974]: *Logical and Epistemological Studies in Contemporary Physics.* Boston Studies in the Philosophy of Science, 13. Dordrecht: D. Reidel Publishing Company. Df. 1.55. Pp. viii+462.

Dillard, A. [1974]: *Pilgrim at Tinker Creek.* New York: Harper's Magazine Press. $7.95. Pp. 271.

Drake, S. [1974]: *Galileo—Two New Sciences.* Madison: University of Wisconsin Press. $14.00 hardback, $5.00 paperback. Pp. vii+323.

Eberlein, G. L. and Kroeber-Riel, W. [1974]: *Wissenschaftstheorie der Wirtschafts Sozialwissenschaften.* Dusseldorf: Verlagsgruppe Bertelsman GmbH. Pp. 309.

Elkana, Y. [1974]: *The Discovery of the Conservation of Energy.* London: Hutchinson Educational Ltd. £4.00. Pp. viii+213.

de Finetti, B. [1974]: *Theory of Probability.* Chichester: John Wiley & Sons. £7.50. Pp. vii+300.

Glucksmann, M. [1974]: *Structuralist Analysis in Contemporary Social Thought.* London: Routledge & Kegan Paul. £4.50. Pp. xii+197.

Goddard, L. and Routley, R. [1974]: *The Logic of Significance and Context,* 1. Edinburgh: Scottish Academic Press. £11.00. Pp. vi+641.

Grossman, R. [1974]: *Meinong.* London: Routledge & Kegan Paul. £4.95. Pp. v+262.

Handy, R. and Harwood, E. C. [1974]: *A Current Appraisal of the Behavioural Sciences.* Great Barrington, Massachusetts: Behavioural Research Council. $12.50. Pp. ii+148.

Hegenberg, L. [1974]: *Definicoes: termos teoricos e significado.* Sao Paulo: Editora da Universidade de Sao Paulo. Pp. 136.

Hesse, M. [1974]: *The Structure of Scientific Inference.* London: Macmillan. £5.95. Pp. vii+309.

Hunter, J. F. M. [1974]: *Essays After Wittgenstein.* London George Allen & Unwin Ltd. £7.50. Pp. vi+202.

Jaki, S. L. [1974]: *Science and Creation—From eternal cycles to an oscillating universe.* Edinburgh: Scottish Academic Press. £4.50. Pp. viii+367.

Jaspers, K. [1973]: *Allegemeine Psychopathologie.* Heidelberg: Springer-Verlag. Pp. iv+748.

Kholeif, F. [1974]: *Avincenna on Psychology: A Study of his Poem on the Soul.* Beirut: Beirut University.

Landmann, M. [1974]: *Philosophical Anthropology.* Philadelphia: The Westminster Press. \$7.50. Pp. 256.

Lenzen, W. [1974]: *Theorien der Bestätigung wissenschaftlicher Hypothesen.* Stuttgart-Bad: Frommann-Holzboog. Pp. 217.

Lessnoff, M. [1974]: *The Structure of Social Science.* London: George Allen & Unwin Ltd. £3.60 hardback, £1.85 paperback. Pp. 173.

Levi, I. [1973]: *Gambling with Truth.* Cambridge, Massachusetts: The M.I.T. Press. \$3.95. Pp. viii+246.

Longuet-Higgins, C. [1974]: *The Development of Mind.* Edinburgh: Edinburgh University Press. £2.25. Pp. 151.

Macbeth, N. [1974]: *Darwin Retried.* London: Garnstone Press. £2.80. Pp. 172.

Madsen, K. B. [1974]: *Modern Theories of Motivation.* Copenhagen: Munksgaard D.Kr 180.00. Pp. 472.

Mann-Paterson, A. [1974]: *Francis Bacon and Socialised Sciences.* Springfield, Illinois: Charles C. Thomas. Pp. vii+191.

Martin, M. [1972]: *Concepts of Science Education.* London: Scott, Foreman and Co. Pp. 176.

Merdeau-Ponty, J. [1974]: *Leçons sur la Genese des Theories Physiques.* Paris: Librairie Philosophique, J. Vrin. Pp. 172.

Merton, R. K. [1973]: *The Sociology of Science.* Chicago and London: The University of Chicago Press. £5.65. Pp. xxxi+603.

Miner, M. [1973]: *General Laws of Behaviour.* Sydney. \$a. 2.00. Pp. 24.

Mischel, T. [1974]: *Understanding Other Persons.* Oxford: Basil Blackwell. £4.00. Pp. vi+266.

McKay, R. and Mitcham, C. [1973]: *Bibliography of the Philosophy of Technology.* Chicago and London: University of Chicago Press. £4.00. Pp. vii+205.

Mittelstrass, J. [1974]: *Die Möglichkeit von Wissenschaft.* Frankfurt: Suhrkamp Verlag. Pp. 268.

Poirier, F. E. [1974]: *In Search of Ourselves.* Minnesota: Burgess Publishing Co. \$4.95. Pp. 400.

Polten, E. P. [1973]: *Critique of the Psycho-Physical Identity Theory.* The Hague: Mouton & Co. D.Fl. 34 Pp. xviii+290.

Porkert, M. [1974]: *The Theoretical Foundations of Chinese Medicine.* Cambridge, Massachusetts: The M.I.T. Press. \$13.95. Pp. xii+368.

Pospesel, H. [1974]: *Introduction to Logic: Propositional Logic.* New Jersey: Prentice-Hall Inc. Pp. v+211.

Quine, W. V. [1974]: *Methods of Logic.* Revised Edition. London: Routledge & Kegan Paul. £3.25. Pp. v+280.

Rabinovitch, N. M. [1973]: *Probability and Statistical Inference in Ancient and Medieval Jewish Literature.* Toronto and London: University of Toronto Press. £6.00. Pp. 205.

Radnitzky, G. [1974]: *The Human Context.* London: Chaucer Publishing Co. Ltd. 60p. Pp. 63.

Schank, R. C. and Colby, K. M. [1974]: *Computer Models of Thought and Language.* San Francisco, California: W. H. Freeman and Co. \$13.50. Pp. vii+454.

Skyums, B. [1966]: *Choice and Chance.* San Francisco, California: Dickenson Publishing Co. Inc. Pp. vii+165.

Sicha, J. [1974]: *A Metaphysics of Elementary Mathematics.* Amherst: The University of Massachusetts Press. \$7.50. Pp. v+444.

Sprigge, T. L. S. [1974]: *Santayana—An Examination of his Philosophy.* London: Routledge & Kegan Paul. £4.50. Pp. v+247.

Stachowiak, H. [1973]: *Allgemeine Modelltheorie.* Vienna: Springer-Verlag. \$45.50. Pp. xv+494.

Stove, D. C. [1973]: *Probability and Hume's Inductive Scepticism.* Oxford: Clarendon Press. £3.00. Pp. 132.

Suppe, F. [1974]: *The Structure of Scientific Theories.* Chicago and London: University of Illinois Press. Pp. viii+682.

Suppes, P. (*ed.*) [1974]: *Space, Time and Geometry.* Dortrecht: D. Reidel Publishing Co. D.fl. 50. Pp. xi+424.

Truitt, W. H. and Solomons, T. W. G. [1974]: *Science Technology and Freedom.* Boston: Houghton Mifflin Company. \$4.50. Pp. viii+272.

TURNER, F. M. [1974]: *Between Science and Religion.* London: Yale University Press. £5.60. Pp. x+273.
WIDMAR, B. [1974]: *La Memoria.* Napoli: Edizioni Glaux. L. 3000. Pp. 183.
WITTGENSTEIN, L. [1974]: *Philosophical Grammar.* Oxford: Basil Blackwell. £7.00. Pp. 495.
ZEIDAN, M. F. [1974]: *Philosophical Method.* Beirut: Beirut Arab University.

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

PROGRAMME OF MEETINGS 1974–75

1974

7 October
Professor N. KOERTGE: Galileo's Philosophy of Science

4 November
Mr W. NEWTON-SMITH: The Topological Structure of Time

9 December
Professor E. GELLNER: The Legitimation of Belief

1975

20 January
Professor T. HUTCHISON: Economics as a Science: Some Present Discontents

17 February
Professor P. RATTANSI: Newton, Man and Myth

17 March
4.15 pm: ANNUAL GENERAL MEETING
5.15 pm: Professor S. KÖRNER: On Cognitive and Practical Rationality

28 April
Mr A. RYAN: 'Rational Man' Explanations in Social Science

19 May
Dr Z. MEDVEDEV: What Future for Research on Ageing and Longevity?

16 June
Professor A. FINE: Einstein on the Quantum Theory

26–28 September
Annual Conference at the University of Warwick

Meetings are held in the Haldane Room, University College, London. Meetings begin at 5.15 pm. Tea is served from 4.45 pm.

THE CENTRE FOR THE PHILOSOPHY OF SCIENCE AND THE DEPARTMENT OF HISTORY AND PHILOSOPHY OF SCIENCE AT THE UNIVERSITY OF PITTSBURGH

An Alternative to the Classical Theories of Perception
JAMES J. GIBSON, Department of Psychology, Cornell University
Tuesday, 1 October 1974

Is Biology Different from Physics?
MICHAEL RUSE, Department of Philosophy, University of Guelph, Ontario
Tuesday, 22 October 1974

Causal Analysis of Non-Experimental Data
PAUL LAZARSFELD, Department of Sociology, University of Pittsburgh
Wednesday, 20 November 1974

Explanations and Final Causes
PETER K. MACHAMER, Department of Philosophy, Ohio State University
Tuesday, 3 December 1974

The Philosophy of Research and the Social Psychology of Research
IAN MITROFF, Graduate School of Business and the Interdisciplinary Program in Information Science, University of Pittsburgh
Wednesday, 15 January 1975

Complementarity and Instrumentalism
MICHAEL R. GARDNER, Department of Philosophy, University of Maryland, College Park
Tuesday, 4 February 1975

Physics and Evidence
CLARK GLYMOUR, Department of Philosophy, Princeton University
Wednesday, 26 February 1975

I *Direct Inference from Statistical Hypotheses*, and
II *Knowledge: Fallibility vs. Corrigibility*
ISAAC LEVI, Department of Philosophy, Columbia University
Tuesday, 11 March 1975
Wednesday, 12 March 1975

History of Science as Philosophy of Science: Towards an Historical Epistemology of Science
MARX W. WARTOFSKY, Department of Philosophy, Boston University
Tuesday, 25 March 1975

All the meetings at G–8 Cathedral of Learning, 8.15 pm. Inquiries can be made by writing to the Director of the Center, Professor Adolf Grünbaum, 2509 Cathedral of Learning, University of Pittsburg, Pittsburgh, PA 15260.

IMRE LAKATOS APPEAL FUND

Many of Imre Lakatos's friends have said that they want to contribute to some lasting memorial to this exceptional man; and there have been many enquiries about the possibility of publishing his unpublished papers. In response to this, an appeal to create an 'Imre Lakatos Appeal Fund' has been launched.

The object of the Appeal Fund will be to promote study in his fields (critical philosophy, the philosophy and history of science and of mathematics, mathematical logic, and related subjects) in various ways. It is hoped in particular to finance the editing and eventual publication of the 'Collected Papers' of Imre Lakatos.

Contributions are invited. Cheques should be made payable to The Imre Lakatos Appeal Fund and should be sent to The Accountant, London School of Economics, Houghton Street, London, WC2A 2AE.

Details of the Fund, together with an account of unpublished manuscripts of Imre Lakatos are available from The Secretary, Department of Philosophy, London School of Economics, Houghton Street, London, WC2A 2AE.

A NEW JOURNAL

A new journal, *Teaching Philosophy*, will appear twice yearly beginning January 1975. The editors, Arnold Wilson and William Todd, invite papers concerned with any facet of the teaching of philosophy and the planning of philosophy courses and curricula. Reports on innovative efforts, articles on the training of teachers of philosophy, and assessments of the likely impact of new teaching ideas for the discipline, as well as notes and news items, are all welcome.

A regular feature will be critical notices of textbooks and surveys of materials available for teaching particular areas – recent texts in introductory logic, for example. Persons interested in contributing such notices, should write the editors, indicating their areas of interest.

Address all communications to The Editors, *Teaching Philosophy*, Mail Location 47, University of Cincinnati, Cincinnati, OH 45221, U.S.A.

Noûs

Nihil philosophici a nobis alienum putamus

A Quarterly Journal of Philosophy

edited by Hector-Neri Castañeda

Philosophy Department, Indiana University

Bloomington, Indiana 47401

NOÛS publishes quality essays and brief discussions on philosophical problems regardless of the author's philosophical school or point of view. From time to time, NOÛS will publish symposia, surveys of work recently done on specially selected philosophical topics, critical studies of recent books (rather than reviews), and catalogues of recent publications in philosophy. NOÛS publishes the symposium papers of the A.P.A. Western Division meetings in its March issue.

A quarterly, NOÛS is published in March, May, September, and November each year. For *individuals* the subscription rates are: one year, $9.00; two years, $17.00; three years, $25.00. For *bona fide students:* $8.00 per year. For *institutions:* $16.00 per year (plus 10% for subscriptions handled through a subscription agency). Add 50 cents per year for all non-U.S. subscriptions. Make checks for one-year subscriptions payable to INDIANA UNIVERSITY NOÛS. Checks for multiple year subscriptions should be made payable to THE INDIANA UNIVERSITY FOUNDATION-NOÛS.

Volume VIII Number 4 November 1974

Tyler Burge—*Truth and Singular Terms*

Laurence F. Mucciolo—*The Identity Thesis and Neuropsychology*

Alan Reeves—*On Truth and Meaning*

T. S. Weston—*Theories Whose Quantification Cannot Be Substitutional*

Paul Teller—*On Why-Questions*

Critical Study

Hector-Neri Castañeda—*Leibniz's Concepts and Their Coincidence* Salva Veritate

Recent Publications in Philosophy

Order from

The Secretary, NOÛS

Department of Philosophy

Indiana University

Bloomington, Indiana 47401

16 APR 1975

MIND

A Quarterly Review of Philosophy

edited by Professor **DAVID HAMLYN**

Vol. LXXXIV JANUARY 1975 No. 333

Articles	
The Argument for Motivation	F. SNARE
Numerical Identity and Objecthood	B. ENC
Distribution Matters	F. SOMMERS
The Always-Painfree Pain-Behaver	R. MONTAGUE
The Use of the Basic Proposition of a Theory of Justice	T. HONDERICH
Science, the Truth, and Thomas Kuhn	H. MEYNELL
Discussions	
Foreknowledge and Predestination Re-examined	J. O'CONNOR
A Rejoinder to O'Connor	L. C. BECKER
The Impermissibility of 'Sufficiency Definitions'	J. J. SHAFER
On the Logic of the Circularity of Logic	G. B. KEENE
Criteriological Arguments in Perception	A. H. GOLDMAN
Cioffi on Freud	B. JONES
Zellner and Imperative Inferences	C. ROBERTS
Quine on Analyticity in L	M. E. LEVIN
Is Light the Proper Object of Vision?	F. R. PICKERING
Critical Notice	
Logic Matters by P. T. Geach	M. CLARK
Book Reviews	
Books Received	

Annual Subscription £2.50 Single Issue 80p

Published for MIND ASSOCIATION by

BASIL BLACKWELL 49 BROAD STREET OXFORD

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY

Edited by MARVIN FARBER

DECEMBER, 1974, Volume XXXV, No. 2

CONTENTS

I. ARTICLES
The A Priori-A Posteriori Distinction, by DAVID W. BENFIELD
The Identity Theory and Criteria for the Mental, by LAURENCE F. MUCCIOLO
Philosophy and Antiphilosophy, by NIKOLAI IRIBADJAKOV
Propositional and Nonpropositional Perceiving, by DAN D. CRAWFORD
On Time and Other Minds, by MELVILLE STRATTON
Brentano, Descartes, and Hume on Awareness, by RICHARD E. AQUILA
II. DISCUSSIONS
III. REVIEWS
IV. NOTES AND NEWS, RECENT PUBLICATIONS

Subscription prices $8.00 per year for individuals; $10.00 per year for institutions. Separate issues, $2,00 and $2.50 respectively. Postage and handling for back issues only. Papers submitted for publication will not be returned unless accompanied by a self-addressed stamped envelope, or equivalent postage.

A Circular listing the main contents of the Journal since 1940 will be sent upon request.

Address all communications to

PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH

State University of New York at Buffalo,
Buffalo, New York 14226

Brit. J. Phil. Sci. Vol. 25 No. 4 December 1974

The British Journal for the Philosophy of Science

Contents

Articles

301–318 John C. Graves Uniformity and Induction

Discussions

319–328 Erhard Scheibe Popper and Quantum Logic

329–332 D. K. Nartonis Quantum Fact and Fiction

332–334 David L. Hull Are the 'Members' of Biological Species 'Similar' to Each Other?

334–336 L. Jonathan Cohen Professor Hull and the Evolution of Science

336–340 S. J. Prokhovnik Did Einstein's Programme Supersede Lorentz's?

340–342 I. J. Good A Little Learning Can Be Dangerous

Review Articles

343–352 J. W. N. Watkins Otto Neurath (review of O. Neurath: *Empiricism and Sociology*);

352–358 M. L. G. Redhead Orthodoxy in Quantum Mechanics (review of E. Scheibe: *The Logic of Quantum Mechanics*);

358–375 Henry E. Kyberg Propensities and Probabilities (review of current literature on the propensity interpretation of probability);

375–397 Howard Stein Maurice Clavelin on Galileo's Natural Philosophy (review of M. Clavelin: *La Philosophie Naturelle de Galilée*)

Recent Publications

Programme of Meetings

Announcements

Printed in Great Britain at the University Press, Aberdeen

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1974-5

President: Professor J. W. N. Watkins
President-Elect: Professor D. J. O'Connor
Hon. Secretary: Helena Sheiham
Hon. Treasurer: Dr D. A. Gillies
Members: Dr M. Machover, Mr H. Martins, Mr N. Maxwell, Dr E. Zahar

EDITORIAL BOARD OF THE JOURNAL

Professor R. B. Braithwaite
Professor H. Dingle
Professor Sir John Eccles
Professor Sir Harold Jeffreys
Professor W. C. Kneale
Professor S. Körner
Professor M. Polanyi
Professor Sir Karl Popper
Professor G. J. Whitrow
Professor J. H. Woodger

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules :

'The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all persons who have been proposed by two members, approved by the Committee, and elected at an ordinary meeting. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £1.00 ($2.50). Those who wish to join should write to the Hon. Secretary, Helena Sheiham, Department of Philosophy, London School of Economics, Houghton Street, London WC2A 2AE.